



Durham E-Theses

Systematic Measurement of Centralized Online Reputation Systems

LIU, LING

How to cite:

LIU, LING (2011) *Systematic Measurement of Centralized Online Reputation Systems*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/881/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Systematic Measurement of Centralized Online Reputation Systems

Ling Liu

A Thesis presented for the degree of
Doctor of Philosophy



School of Engineering and Computing Sciences
Durham University

April 2011

Abstract

Background: Centralized online reputation systems, which collect users' opinions on products, transactions and events as reputation information then aggregate and publish it, have been widely adopted by Internet companies. These systems can help users build trust, reduce information asymmetry and filter information.

Aim: Much research in the area has focused on analyzing single type systems and the cross-type evaluation usually concentrates on one aspect of the system. This research proposes a systematic evaluation model (SERS) that can measure different types of reputation system.

Method: From system perspective, all reputation systems can be divided into five underlying components. *Input* refers to the collection of ratings and reviews; *Processing* is the aggregation of ratings. *Output* publishes the information. *Feedback Loop* is the collection of the feedback of the review, which can be seen as the 'review of the review'; Finally, *Storage* stores all the information. Therefore, based on each component's characteristics, a series of benchmark criteria can be defined and incorporated into the model.

Results: The SERS has defined 29 criteria, which can compare and measure different aspects of reputation systems. The model was theoretically assessed on its coverage of the successful factors of reputation systems and the technical dimensions of information systems. The model has also been empirically assessed by applying it to 15 commercial sites.

Conclusion: The results obtained indicated that the SERS model has identified most important characteristics that have been proposed by reputation systems literature. In addition the SERS has covered most dimensions of the two basic technical information system measurements: information quality and system quality. The empirical assessment has shown that the SERS can evaluate different types of reputation systems and is capable of identifying the weakness of current systems.

Declaration

The work in this thesis is based on research carried out at the Innovative Computing Group, the School of Engineering and Computing Sciences, Durham University. No part of this thesis has been submitted elsewhere for any other degree or qualification and it all my own work unless referenced to the contrary in the text.

Publications:

- Liu L, Munro M, Song W (2010). Evaluation of collecting reviews in centralized online reputation systems. In: *the 6th International Conference on Web Information Systems and Technologies (WEBIST)*. Valencia, Spain.
- Liu L and Munro M (under review). ‘The Pursuit of Quality Outputs – The Assessment of the Output of Centralized Reputation Systems’. Submitted to *Journal of Information Science*.
- Liu L and Munro M (under review). ‘Systematic Evaluation of Centralized Online Reputation Systems’. Submitted to *Decision Support Systems*.

Copyright © 2011 by Ling Liu.

“The copyright of this thesis rests with the author. No quotations from it should be published without the author’s prior written consent and information derived from it should be acknowledged”.

Acknowledgements

First of all, I owe my deepest gratitude to my supervisors, Professor Malcolm Munro and Dr. William Song for the excellent supervision and guidance they have both given me. I have learnt so much throughout my PhD and without their invaluable advice and encouragement, this thesis would simply not exist.

I am indebted to those who helped to make this thesis possible. Thanks to Professor David Budgen who has given me precious advices on the research methodologies and I am also grateful for M. Imran who made the thesis latex template.

Finally, I would like to thank my parents and my husband for their continued love and for understanding and respecting my need to work. I could not have finished without their support.

Contents

| | |
|--|------------|
| Abstract | ii |
| Declaration | iii |
| Acknowledgements | iv |
| 1 Introduction | 1 |
| 1.1 Background | 1 |
| 1.2 Criteria for Success | 3 |
| 1.3 Structure of the Thesis | 4 |
| 2 Literature Review | 6 |
| 2.1 Introduction | 6 |
| 2.2 The Notion of Reputation | 6 |
| 2.3 Online Reputation Systems | 8 |
| 2.3.1 Trust Building | 9 |
| 2.3.2 Reducing Information Asymmetry | 12 |
| 2.3.3 Information Filtering | 14 |
| 2.3.4 Classifications of Reputation System | 16 |
| 2.4 Evaluation of Reputation Systems | 18 |
| 2.4.1 C2C Systems | 18 |
| 2.4.2 Review Centres | 20 |
| 2.4.3 Online Communities | 21 |
| 2.4.4 Cross-type Evaluation | 22 |
| 2.4.5 Successful Factors of Reputation Systems | 23 |

| | | |
|----------|--|-----------|
| 2.4.6 | Summary | 26 |
| 2.5 | Evaluation of Information Systems | 26 |
| 2.5.1 | Information Quality (IQ) | 29 |
| 2.5.2 | System Quality (SQ) | 31 |
| 2.5.3 | Summary | 32 |
| 2.6 | Conclusion | 33 |
| 3 | Reputation System Terminology and Structure | 35 |
| 3.1 | Introduction | 35 |
| 3.2 | Essential Entities | 35 |
| 3.3 | The Structure of Reputation Systems | 36 |
| 3.3.1 | Input | 38 |
| 3.3.2 | Processing | 39 |
| 3.3.3 | Output | 40 |
| 3.3.4 | Feedback Loop | 40 |
| 3.3.5 | Storage | 41 |
| 3.4 | Conclusion | 41 |
| 4 | The SERS Model | 44 |
| 4.1 | Introduction | 44 |
| 4.2 | Input | 45 |
| 4.2.1 | Collection Channel | 45 |
| 4.2.2 | Information Sources | 45 |
| 4.2.3 | Reputation information | 51 |
| 4.2.4 | Collection Costs | 53 |
| 4.3 | Processing | 59 |
| 4.3.1 | Algorithms | 59 |
| 4.3.2 | Attributes of Algorithms | 63 |
| 4.4 | Output | 64 |
| 4.4.1 | Dissemination | 65 |
| 4.4.2 | Aggregated Information | 66 |
| 4.4.3 | Individual Information | 67 |

| | | |
|----------|--|-----------|
| 4.4.4 | Response Time | 68 |
| 4.5 | Feedback Loop | 69 |
| 4.5.1 | Feedback Function | 69 |
| 4.5.2 | Feedback Collection | 70 |
| 4.5.3 | Feedback Loop Collection Cost | 72 |
| 4.6 | Storage | 73 |
| 4.6.1 | Input Information Storage Cost | 73 |
| 4.6.2 | Feedback Storage Costs | 75 |
| 4.6.3 | Processing Information Storage Costs | 76 |
| 4.7 | Conclusion | 77 |
| 5 | Theoretical Evaluation of the SERS | 79 |
| 5.1 | Introduction | 79 |
| 5.2 | Influential Factors | 79 |
| 5.3 | IS Dimensions VS the SERS Criteria | 80 |
| 5.3.1 | IQ Dimensions | 81 |
| 5.3.2 | SQ Dimensions | 83 |
| 5.3.3 | Summary | 85 |
| 5.4 | Conclusion | 85 |
| 6 | Empirical Evaluation of the SERS | 87 |
| 6.1 | Introduction | 87 |
| 6.2 | Research Design | 87 |
| 6.2.1 | Commercial Sites | 87 |
| 6.2.2 | Data Collection | 91 |
| 6.3 | Results | 92 |
| 6.3.1 | Input | 92 |
| 6.3.2 | Processing | 103 |
| 6.3.3 | Output | 107 |
| 6.3.4 | Feedback Loop | 115 |
| 6.3.5 | Storage | 120 |
| 6.4 | Conclusion | 124 |

| | |
|------------------------------------|------------|
| 7 Conclusion | 128 |
| 7.1 Introduction | 128 |
| 7.2 Contributions | 129 |
| 7.3 Criteria for Success | 130 |
| 7.4 Limitations | 131 |
| 7.5 Future Work | 132 |
| References | 134 |

List of Figures

| | | |
|-----|--|----|
| 2.1 | Reputation Typology (Mui et al., 2002b) | 7 |
| 2.2 | eBay’s Feedback Profile | 11 |
| 2.3 | Amazon’s Product Review | 14 |
| 2.4 | Snapshots of Digg | 15 |
| 2.5 | Classifications of Reputation System | 17 |
| 2.6 | Classification by the Functions | 17 |
| 2.7 | ‘Helpful’ Vote and Spam Report | 21 |
| 2.8 | The D&M IS Success Model (Delone and McLean, 2003) | 28 |
| 3.1 | Essential Entities of Reputation System | 35 |
| 3.2 | Different Structure Models | 36 |
| 3.3 | Reputation System Structure Model | 37 |
| 3.4 | The Use Case Diagram of Evaluator and Information Source | 38 |
| 3.5 | The Use Case Diagram of EC Provider | 39 |
| 3.6 | The Use Case Diagram of End User | 40 |
| 3.7 | The Use Case Diagram of Feedback Provider | 41 |
| 3.8 | The Use Case Diagram of System Users | 42 |
| 3.9 | The Activity Diagram of Reputation System | 43 |
| 4.1 | The Sequence Diagram of Direct Channels | 46 |
| 4.2 | The Sets of U_q | 47 |
| 4.3 | Snapshot of Tripadvisor ‘Write a Review’ page | 52 |
| 4.4 | The Activity Diagram of C_{1a} System | 54 |
| 4.5 | The Activity Diagram of C_{1b} System | 57 |
| 4.6 | C_2 System Collection Process | 58 |

| | | |
|------|--|-----|
| 4.7 | Output Criteria | 65 |
| 4.8 | Ratings Presented in Different Time Periods | 67 |
| 4.9 | Different Functions of Feedback Loop | 70 |
| 4.10 | Reddit's Multi-Level Feedback | 72 |
| 6.1 | Distribution of Collection Channels | 93 |
| 6.2 | Comparison of the Set of Evaluators | 94 |
| 6.3 | Comparison of the Granularity of Evaluators | 95 |
| 6.4 | Information Breadth and Format | 97 |
| 6.5 | Input Collection Costs (Min. and Max. Settings) | 101 |
| 6.6 | Input Collection Costs with 150-words | 102 |
| 6.7 | The Numbers of Ratings and Text Reviews of the Same Target | 103 |
| 6.8 | System Complexity Comparison | 107 |
| 6.9 | Different Information Presentation from Yahoo! Answers | 107 |
| 6.10 | Dissemination Methods | 108 |
| 6.11 | Timeliness Distribution | 109 |
| 6.12 | Descriptive Dimensions | 110 |
| 6.13 | Information Sorting and Filtering Comparison | 111 |
| 6.14 | Feedback Function | 116 |
| 6.15 | Sets of Feedback Providers | 117 |
| 6.16 | Feedback Breadth | 117 |
| 6.17 | The Nature of Rating Feedback | 118 |
| 6.18 | Feedback Level | 119 |
| 6.19 | Feedback Collection Cost | 120 |
| 6.20 | The Value of S_{pr} (Bytes) | 121 |
| 6.21 | The Value of S_{pr} (Bytes) | 122 |
| 6.22 | The Comparison of S_{fd} | 123 |

List of Tables

| | | |
|-----|--|-----|
| 2.1 | Successful Factors of Reputation System | 27 |
| 2.2 | IQ and SQ Dimensions | 33 |
| 4.1 | Criteria Summary | 78 |
| 5.1 | The SERS vs the Successful Factors of Reputation Systems | 80 |
| 5.2 | IQ and SQ Dimensions | 81 |
| 5.3 | IS Dimensions and the SERS | 86 |
| 6.1 | Summary of Selected Sites | 89 |
| 6.2 | Sites with Special Limitations | 100 |
| 6.3 | P1, P2 and P3 Algorithms | 103 |
| 6.4 | Evaluator Information | 113 |
| 6.5 | Feedback Loop Information | 114 |

Chapter 1

Introduction

1.1 Background

Since the middle of 1990s the Internet has become a very important part of our lives. People use the Internet every day to send emails, chat with friends, listen to music, play games and most importantly seek information (Dutton et al., 2005). The results in the First World Internet Report (Pierce, 2008) showed that more than two thirds of internet users consider the Internet as a very important source of information. However it is not easy to find desired information in this information age — the Internet is too big. Google, the most used search engine, announced that by July 2008, it has indexed of 1 trillion unique URLs, compared to when Google first founded in 1998, it was only 26 million pages ¹.

The biggest problem information overload brought is the difficulties to make decisions. Individuals need sufficient information to make decisions, the more information they gain the better decision they can make — up to a certain point (Chewning and Harrell, 1990). After the point, more information will confuse the individual and affect their ability to set priorities (O'Reilly III, 1980; Chewning and Harrell, 1990; Eppler and Mengis, 2004).

Online Reputation Systems are one of the best ways to solve the problem. *Online*

¹Alpert, J. and Hajaj, N. (2008). 'We knew the web was big...'. <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>; Last Accessed 15 January 2011.

Reputation Systems use internet technologies to build large-scale word-of-mouth networks from former users' experience and aggregate information to derive a trust or reputation score, which can assist prospective users to make decisions (Dellarocas, 2003).

Based on their information storage location, reputation systems can be divided into two main types. *Centralized Reputation Systems* rely on a central entity to gather, compute and disseminate reputation information. Centralized reputation systems are widely used in the following areas: Consumer-to-Consumer (C2C) markets, online retailers, shopping comparison sites and information communities.

Distributed Reputation Systems on the other hand rely on decentralized solutions where every peer stores information about the other peers with which they interacted. Reputation information is disseminated on demand between peers. Distributed systems are mainly used within peer-to-peer systems (Jøsang et al., 2007).

This research focuses on the Centralized Reputation Systems only. Therefore, in the rest of the thesis, all reputation systems refer to the centralized ones unless otherwise noted.

Since the end of the 20th century, reputation systems have been widely adopted by Internet companies and they naturally have different interfaces and track different aspects of user behavior (Friedman et al., 2007). For example, to build trust between strangers, eBay.com, one of the largest marketplaces on the Internet, allows their buyers and sellers to leave positive, neutral or negative feedback on each other. Amazon.com, the largest online retailer, encourages their users to write reviews on their products so that potential consumers can gather more information about the products (David and Pinch, 2005; Dellarocas, 2003). Furthermore, by taking advantages of 'the wisdom of the crowd' (Surowiecki, 2005), reputation systems can be used to filter information. Digg.com is a website that allows people to share internet content by submitting links of the stories. Voting stories up ('digging') and down ('burying') is the site's cornerstone function. Each story has a number associated with it, which is calculated by the number of 'diggs' minus the number of 'buries'. Larger numbers indicate more interesting stories in the opinion of the readership.

Reputation systems have different interfaces, functions and types. However much research in the area has focused on single type of system rather than comparing different systems together. Thus, this research proposes an evaluation model, which will be called the SERS model, to measure different kinds of system in the same context. The research focuses on analyzing the intrinsic nature of reputation systems from the structure perspective.

1.2 Criteria for Success

This section describes the criteria for assessing the success of the research.

1. **A model that can represent the major characteristics of reputation system.**

As a comprehensive evaluation model, it should be able to illustrate different aspects of reputation systems. A reputation system is a complicated structure, including many processes and activities. In other words, only assessing one aspect of the system is not sufficient. Therefore, this criterion specifies that the evaluation model should be able to illustrate major characteristics of reputation systems.

2. **The model should consider the cost of reputation systems.**

Generally a high performance of a system often comes with a high cost. Therefore, when assessing a system, it should not only consider how well it performs but also how much it costs. It should be noted, this research concentrates on the intrinsic nature of reputation systems, thus the management costs, such as human resources, hardware costs and most 'money costs', will not be considered. Rather, the model should take more considerations on the time and system costs.

3. **The model can be empirically evaluated using samples taken from the commercial world.**

The model is built on a theoretical level, the best way to validate it is to apply

the model to commercial sites. In other words, the model should be applied to evaluate a number of commercial sites.

4. The model can compare and measure different types of reputation system.

The evaluation model must be able to measure different types of system instead of focusing on only one type. Furthermore, the model should be able to show the differences between them.

These criteria will be revisited and discussed in the final chapter (Chapter 7) of this thesis.

1.3 Structure of the Thesis

The rest of the thesis is organized as follows.

Chapter 2 surveys the literature that surrounds the online reputation systems and related topics. It begins by giving an overview of the concept of reputation; then discusses the nature of online reputation systems, in particular, their functions. Next, the chapter provides a discussion of the literature focusing on the evaluation of reputation systems. Considering that reputation systems are essentially information systems, literatures in information systems evaluation area has also been reviewed.

Chapter 3 is concerned with the terminologies of entities that are related to reputation systems. Following this, an analysis of the structure of reputation systems is provided with emphasis on their underlying components.

Chapter 4 provides a full discussion on the evaluation criteria of the SERS model. It first discusses the characteristics of each component and then defines a series of benchmark criteria accordingly. The chapter also examines the influential factors of each criterion or the possible quantifications. In total, 29 criteria have been defined and grouped into classification criteria, measurement criteria and cost criteria.

Chapter 5 evaluates the SERS from theoretical perspective. It first compares the SERS with the successful factors of reputation systems that are proposed and

discussed by relevant literature. Then the SERS is compared with the quality dimensions of the information system evaluation model. The aim of the chapter is to testify whether the SERS has a good coverage of the most important characteristics and factors of reputation systems.

Chapter 6 evaluates the SERS by applying it to a number of commercial sites, which represent different types of reputation system. The results are supposed to show whether the SERS is able to classify and measure different commercial systems.

Chapter 7 presents the conclusions of the thesis and summaries the research carried out. The recognized limitations of this research are also described. Finally potential further work is suggested.

Chapter 2

Literature Review

2.1 Introduction

This chapter surveys the literature related to the reputation systems. It first discusses the notion of reputation and how people rely on it to make decisions every day. Then Section 2.3 introduces the concept of reputation systems and presents a detailed description of the functions of these systems. Section 2.4 reviews the literatures in reputation systems evaluation area. With the considerations of reputation systems are basically information systems, Section 2.5 discusses relevant literature in the evaluation of information systems area.

2.2 The Notion of Reputation

According to The Oxford English Dictionary (1992), *reputation* is defined as ‘The common or general estimate of a person with respect to character or other qualities; the relative estimation or esteem in which a person or thing is held’ (Volume XIII, Page 678). It is a characteristic or attribute ascribed to one person by another and is the opinion which is publicly formed and held. In ‘Trust in Modern Society’, Misztal pointed out that the construction of a reputation within the community depends on members having sufficient information about one another’s past behavior (Misztal, 1996). From this point of view, reputation is a state used to predict people’s future actions based on their past actions (Wilson, 1985).

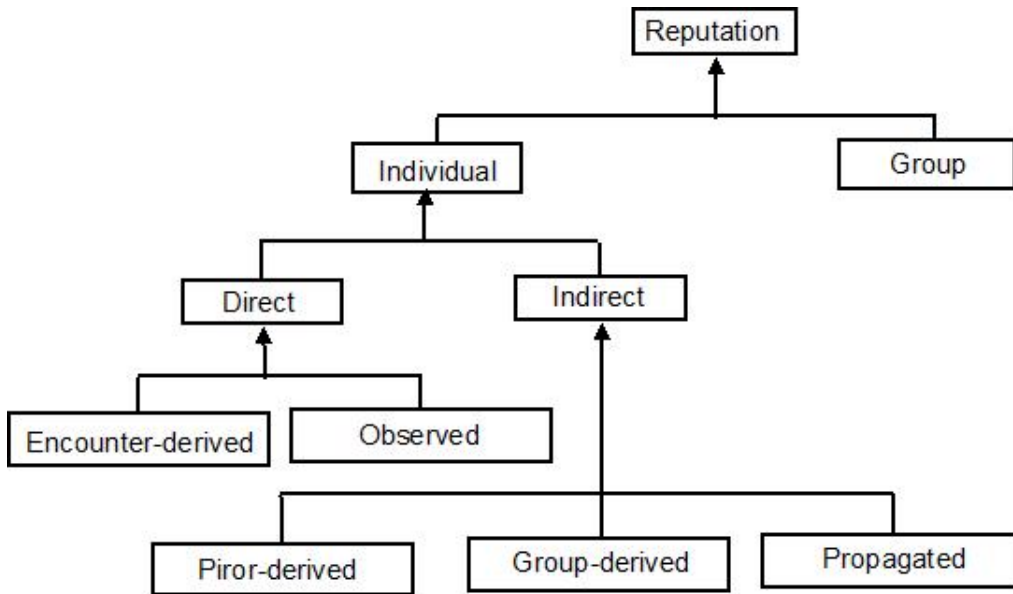


Figure 2.1: Reputation Typology (Mui et al., 2002b)

Mui et al. (2002b) proposes a comprehensive classification of reputation by the means of collecting it (Figure 2.1). At the topmost level, reputation can be used to describe an individual or a group of individuals. The authors consider individual reputation as being derived either from direct encounters or observations or from inferences based on information gathered indirectly. Therefore, there are two main classifications of individual reputation: *Direct Reputation* refers to reputation estimates by an evaluator based on direct experiences, such as, interacting with the other agent or observations made about another agent’s encounters with others. *Indirect Reputation* refers to reputation estimates that are based on secondhand evidence. Without direct evidence, individual reputation can be inferred based on information gathered indirectly: individual’s personal characters (Prior-derived reputation), the social group individual belongs to (Group-derived Reputation) and information gathered from word-of-mouth (Propagated Reputation) (Mui et al., 2002b).

People make decisions depend on reputation everyday. For instance, if someone looks for a plumber, usually they will ask their friends or neighbours for suggestions. Friends and neighbours will then share experiences on their previous hired plumbers. That experiences form the reputation of the plumbers. Similarly, when choosing from two restaurants, people usually select the one has more customers. Because

more people there means the restaurant may provide better food and services. This is reputation.

Although reputation is one of the most important factors that help people make decisions, there are two limitations with traditional reputation information (Dellarocas et al., 2004):

- **Local distribution.** It is not difficult for neighbours and friends to share their experience; however it is difficult to distribute information more widely and efficiently. Therefore, people can only obtain reputation information within a small scale.
- **Impermanency.** In the real world, reputation information is difficult to ‘store’. People talk about it, pass it from one to another. But the dissemination will not last long, which means it is not easy to retrieve past information.

The development of computer and Internet technologies bring reputation management into a new age.

2.3 Online Reputation Systems

One of the biggest advantages that Internet offers is it largely reduced the transaction costs of collecting, processing and disseminating information. It creates new opportunities for people communicating with others and sharing their opinions and experiences out of the local area, which can be extended to a national or even world wide scale. Reputation systems take advantages of the Internet to redefine the age-old concept of word-of-mouth (Dellarocas, 2006a).

Online Reputation Systems collect people’s opinions on products, transactions and events as reputation information (ratings or text reviews), and then aggregate this information and disseminate it to the public, so that other people can use the reputation information as a reference to make decisions (Dellarocas, 2003; Resnick et al., 2000).

Reputation systems are playing a very important role on the Internet, particularly for consumers. The first two purchase influencers in the US are both

from the power of reputation systems: personal advice from friends and online reviews/comments from buyers (Rubicon Consulting, 2008). Similarly, the results in ‘Nielsen Online Holiday Survey 2008’ showed that online reviews already influence consumers ‘offline shopping’ behaviors. According to the report, more than 80% of respondents said they have read product or retailer reviews by other customers before visiting the physical store (Nielsen, 2008).

Academic research has shown similar results. A number of researchers have found that consumers believed that online reviews are more reliable than traditional source of information (Bickart and Schindler, 2001; Clemons, 2008; Koh et al., 2010).

Reputation systems have a positive impact on business organizations as well. The provision of reputation systems can increase performance (Yang et al., 2007), e.g., the usefulness and social presence of the websites (Kumar and Benbasat, 2006). In addition, the adoption of reputation systems can increase the loyalty of the web site (Chen et al., 2009) and customer satisfaction (Morzy, 2008).

Research also found that reputation information can be used to predict sales. Dellarocas et al. (2004) claimed that online movie reviews can be exploited for revenue forecasting and planning by analyzing the relationship between the movie reviews on the Yahoo! web site and movies revenue.

As Dewan and Hsu (2004) pointed out that reputation systems have become an essential part of online auctions, e-storefronts, and a wide-range peer-to-peer systems around the Internet, including leading online companies. Online reputation systems have three main functions: building trust among strangers, reducing information asymmetry and filtering information. The following sections describe these functions in detail.

2.3.1 Trust Building

Trust is vital to human society, people experience and rely on it everyday. According to The Oxford English Dictionary (1992), trust is defined as ‘Confidence in or reliance on some quality or attribute of a person or thing, or the truth of a statement’(Volume XVIII, P. 623). The definition pointed out the two aspects of trust: confidence and reliance. Although the notion of trust has been studied

from many disciplines, including psychology, sociology and economics (Wang and Emurian, 2005; Sabater and Sierra, 2005), much research has shown that individuals' perceptions of others' trustworthiness and their willingness to engage in trusting behavior when interacting with them are largely history-dependent processes (Kramer, 1999).

There are two forms of trust (Frowe, 2005): primary trust and secondary trust. *Primary trust* specifies trusts which built on direct personal interactions or observations. *Secondary trust* refers to the tacit trust relationships people have with those individuals or institutions that they do not encounter directly but nevertheless trust to act in certain ways. In other words, secondary trust is built on the individuals or institutions' reputation.

Pettit (2004) argued that in the Internet environment, it is difficult to build primary trust, which means, the trust among strangers will only rely on one another, this makes the 'online trust is fantasy'. However some researchers, for example, Laat (2005) believed that one's reputation could be used to assess their trustworthiness without observing their characteristics. Falcone and Castelfranchi (2001) and Good (1988) also indicated that reputation is one of the most important factors for assessing trust. As long as agents value their esteem, the long-term reputation based trust could be well constructed on the web. Online reputation systems are the best examples.

Take eBay.com as an example, when a buyer purchases an item from eBay, most of the time, they are required to pay the bill before the seller dispatches the goods. After the buyer has paid the bill, the seller has two choices: send the goods or keep it. Therefore, the buyer is facing a risk of losing both money and goods. This is the situation of moral hazard (Dellarocas, 2006b).

Reputation information can deter moral hazard by its *sanctioning role*. As Misztal (1996) indicated the sanctioning role of reputation information forces individuals to recognize that their own behavior has consequences for their reputation and eventually it will influence their own welfare. In other words, reputation information forces entities to keep honest therefore, it can help build trust between strangers.

Figure 2.2 shows a sellers feedback profile on eBay. The site employs a so-called

‘feedback system’ that allows buyers and sellers to leave positive (+1), neutral (‘0’) or negative (-1) feedback to each other. Each buyer/seller then has an overall score, which is the aggregation of all ratings. A lower score indicates a less trustworthy buyer/seller. Short text comments are also allowed as the complement of the ratings. Along with the overall rating, buyers can also leave Detailed Seller Ratings (DSR) to sellers. Unlike the overall rating, detailed seller ratings are collected based on 5 level Likert Scale. Five stars represents the best service, one is the worst. The detailed ratings consist of the performance of the delivery, packaging and communications of seller’s service.

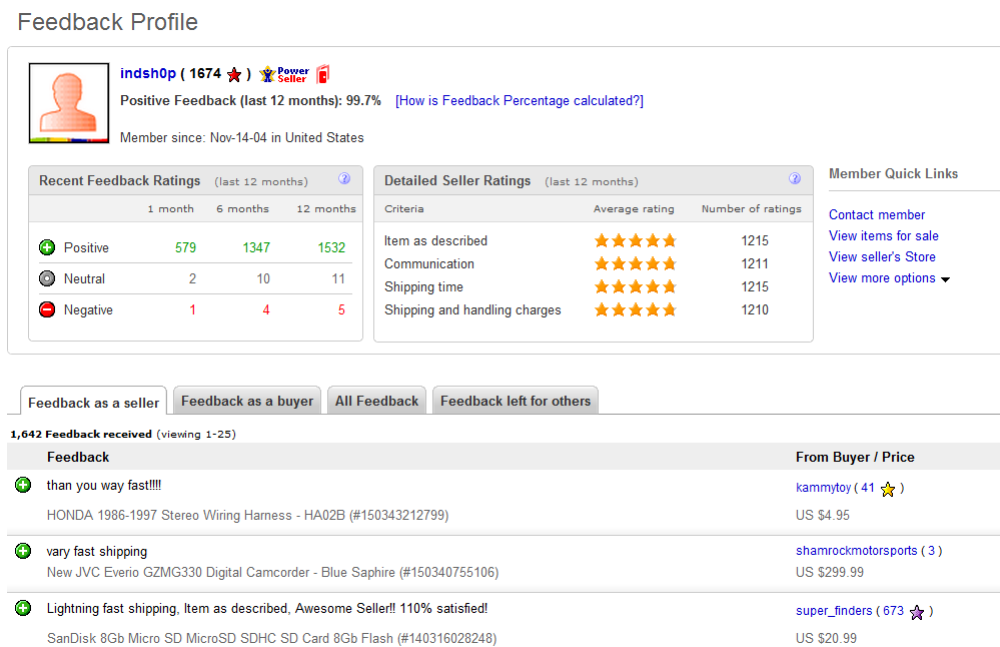


Figure 2.2: eBay’s Feedback Profile

Most consumer-to-consumer (C2C) sites utilize eBay-like reputation systems. Laboratory simulation has shown that these systems can effectively build trust among strangers. Bolton et al. (2004) constructed three markets: 1) *Stranger market* is the market where individual buyers and sellers meet no more than once and the buyer has no information about the seller’s transaction history. 2) *Feedback market* tracks seller histories of shipping decisions and provides this information to prospective buyers. 3) In the *partners market*, the same buyer-seller pairs interact repeatedly. By comparing robustness of the three markets, the authors found that

although the feedback market is less efficient than the partner market, it is more efficient than the stranger market.

Furthermore these systems can also help buyers to avoid online auction fraud. Gregg and Scott (2006) found that recent negative feedback posted in an online reputation system is useful in predicting future online auction fraud and the experienced online auction buyers are in a better position to use reputation information to avoid potentially fraudulent auctions.

In addition, many studies have shown that a seller with a higher reputation score in an online auction site can sell their items for a higher price (Lucking-Reiley et al., 2007; Houser and Wooders, 2006; McDonald and Slawson, 2002; Ba and Pavlou, 2002). Resnick et al. (2006) conducted the first randomized controlled field experiment of the eBay reputation mechanism. The authors observed the results of selling the same goods (vintage postcards) from two identities: that of a new seller and that of a highly reputable seller. As predicted, the seller with a good reputation did significantly better, and obtained, on the average, 8.1% higher prices than the new seller.

2.3.2 Reducing Information Asymmetry

One important premise of a fair market environment is that buyers and sellers in the marketplace are perfectly informed (Klang, 2001). However sometimes, one party has more or better information than the other, this situation is *information asymmetry* (Aboody and Lev, 2000; Healy and Palepu, 2001; Pindyck and Rubinfeld, 2009). Information asymmetry may lead to a situation, where the bad products or customer are more likely to be selected (Pindyck and Rubinfeld, 2009; Dellarocas, 2005). This is called *adverse selection*. Akerlof (1970) takes examples from the second-hand car markets, where sellers know what the buyers cannot judge the true quality and value of the cars. In his model, unless sellers can credibly signal the product quality, buyers are willing to pay only the expected average price. Sellers with higher-quality products are unwilling to sell at the lower average price (Akerlof, 1970). The paper shows that adverse selection will eventually drive all, except the lowest quality sellers, out of the market.

As Friedman et al. (2007) indicated that because histories reveal information about abilities, entities with higher abilities will be drawn to participate, as they will be distinguishable from those of lower abilities, and respected or rewarded appropriately. In other words, visible histories avoid problems of adverse selection. The *signaling role* of reputation information can make histories visible. For example, hotels have more information on their service quality than perspective customers. By publishing experiences of customers who have stayed in the hotel, other consumers can learn the true quality of it.

Word-of-Mouth is an age old concept, which refers to ‘all informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services or their sellers’ (Westbrook, 1987). Traditionally it is believed to have great impact on consumer decision-making (Bone, 1995). The Internet has enhanced its power.

The Internet is considered to be the best way to reduce information asymmetry as it provides opportunities for users to share their reviews and comments on products. Many online retailers and price comparison sites employed reputation systems. As one of the biggest online retailers, Amazon allows users to leave ratings and reviews on all products sold on its website. Users can rate products from 1 to 5, and write a text review with it. Therefore, potential customers can read the reviews and then obtain more information on the products from them. Figure 2.3 shows the product reviews on Amazon.

An increasing number of studies have found a positive relationship between online consumer reviews and sales of products including books, movies, TV shows and video games (Zhu and Zhang, 2010; Godes and Mayzlin, 2004). Zhang and Delarocas (2006) suggested that online movie reviews have a positive and statistically significant influence on other people. By developing a diffusion model analyzing the relationship between ratings from online movie review sites and weekly movie revenues, the authors found that a 1-point increase in an overall rating (5 point rating scale) can induce 4-10% more people to watch the movie. Chevalier and Mayzlin (2006), which focus on the book reviews, also indicated that an improvement in a book’s reviews can lead to an increase in its sales. Cui et al. (2010) pointed out that



Figure 2.3: Amazon's Product Review

negative reviews affects new products sales more than positive reviews.

In addition to the impact on consumers, online reviews are also believed to have positive influences to organizations and online sellers. Chen and Xie (2008) argued that online reviews can serve as a new element in the marketing communications strategies and work as a 'sales assistant'. Dellarocas (2003) indicated that consumer reviews can affect a wide range of activities within organizations, such as brand building and customer acquisition, product development and quality control and supply chain quality assurance.

2.3.3 Information Filtering

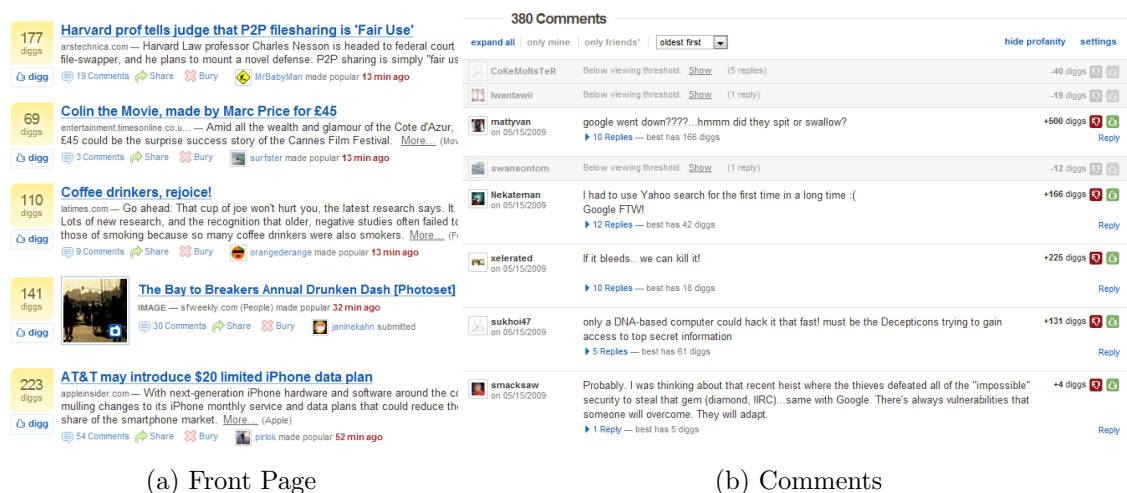
In the information age, one of the most important issues is how to filter information. Unlike products or items, the quality of information is extremely difficult to assess. Goldberg et al. (1992) suggested one possible solution, collaborative filtering. As the authors stated, collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read.

By using the 'wisdom of crowds', the opinion of a group of people can be used to assess the information quality. Surowiecki (2005) argued that under the right circumstances, groups are remarkably intelligent and can make the right judgment.

It worked in academic literature area, in where the quality of an academic publication usually can be assessed by how many times it has been cited.

Similarly, web sites can borrow the idea and let their users decide the quality of the documents they have read, and then it can effectively solve the problems of document recommendation and rating (Lerman, 2006).

As discussed in Section 1.1, Digg allows users to up ('digg') and down ('bury') stories, so that an interesting story can be identified by a larger number. Many stories get submitted every day, but only the most Dugg stories appear on the front page. Although Digg does not disclose their algorithm publicly, they do take several factors into consideration, including (but not limited to) the number and diversity of diggs, buries, the time the story was submitted and the topic ¹. Figure 2.4a is the interface of its homepage. Digg also allows users to leave comments, which can be 'diggged' and 'buried' as well, on articles and stories. Figure 2.4b shows how the comments look like. Some comments are omitted because they have been 'buried' too many times. However, if users want, they may choose to 'show' the omitted ones. The number of 'diggs' and 'burys' can also be retrieved.



(a) Front Page

(b) Comments

Figure 2.4: Snapshots of Digg

Lampe and Resnick (2004) analyzed another similar site, Slashdot.org, a news forum which focused on technology information. The site's editors select a number

¹Digg (2011). 'Digg faq.' <http://digg.com/faq>; Last Accessed 15 January 2011.

of news stories and publish them on the site, usually each story can attract a couple of hundreds of comments. Unlike Digg, their moderation system only allows users to rate the worth of comments. With the results of their analysis, Lampe and Resnick (2004) indicated that although it often takes a long time for good comments to be identified, the moderation mechanism can consistently separate high and low quality comments.

2.3.4 Classifications of Reputation System

Reputation systems can be classified into many different types. As discussed in Section 1.1, based on the network architecture, which determines how information is gathered and stored, reputation systems can be classified into *centralized* and *distributed* systems (Gutowska, 2009; Jøsang et al., 2007).

Cho et al. (2009) divided reputation systems into: *explicit mechanisms* and *implicit mechanisms* on the basis of information source. The former source voluntarily write reviews or provide ratings, whereas the latter derived information from users activities, for example, the best selling books are ranked by the number of sales. Section 3.3.1 has more details on the differences between these two types, although this thesis believes the differences is based on the type of information rather than the information source.

Reputation systems can also be classified depending on their e-business model (Gutowska, 2009; Cho et al., 2009). *Bidirectional systems*, which are mostly used by C2C and Peer-to-Peer (P2P) sites, allow users to rate each other. In *unidirectional systems*, where users give ratings or write reviews on products or services, are mostly adopted by Business-to-Consumer (B2C) companies. However, this classification has ignored the information filtering function of reputation systems. In addition, Digg or Slashdot also use the unidirectional systems, but the sites cannot be simply considered as B2C companies.

Figure 2.5 shows the different classification of reputation systems. The last classification is based on the functions of reputation systems, which have been discussed in the previous sections.

One advantage of classifying systems by their functions is that the function of

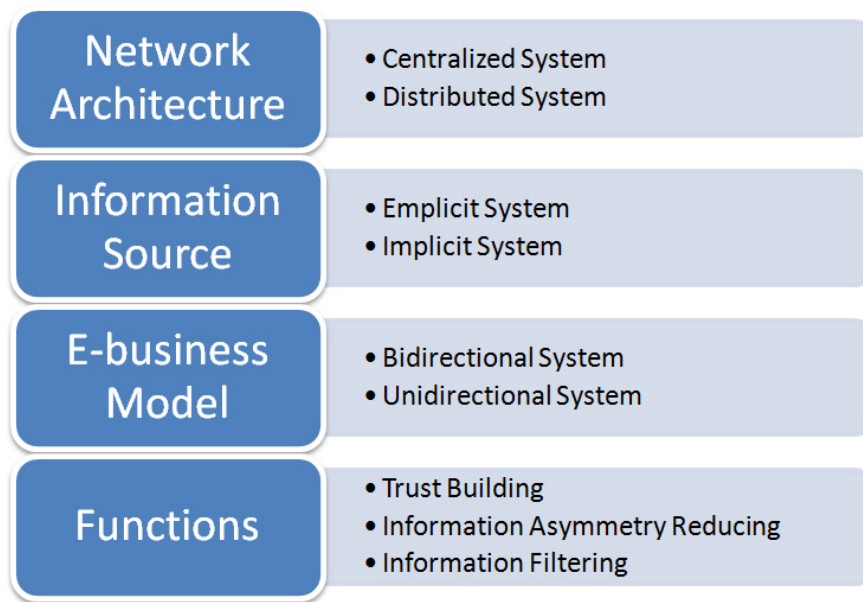


Figure 2.5: Classifications of Reputation System

the system is usually connected with the nature of the web sites. Figure 2.6 shows that most C2C marketplaces use reputation systems for building trust among the buyers and sellers, whereas online retailers, price comparison sites and review centres depend on reputation systems to reduce information asymmetry. Information centres and online forums adopt reputation systems to filter information.

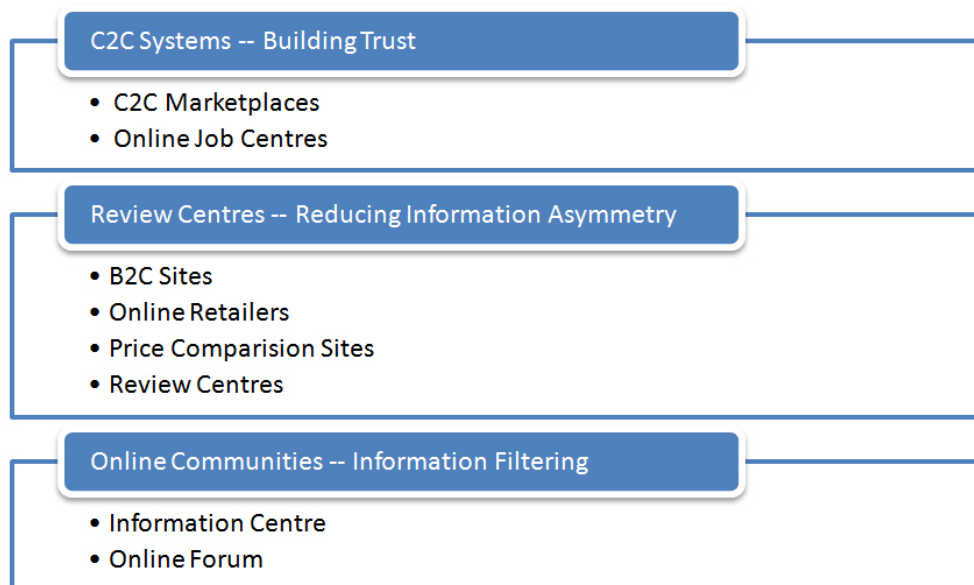


Figure 2.6: Classification by the Functions

The rest of the thesis uses the function classification for further analysis and discussion. For the sake of convenience, *C2C Systems*, *Review Centres* and *Online Communities* refer to the different types respectively.

2.4 Evaluation of Reputation Systems

This section reviews the literature related to the evaluation of reputation systems. The majority of studies in the area focused on assessing single type systems, in particular C2C Systems and Review Centres, from economics and other social science perspective. Very few studies compare different types of system together.

2.4.1 C2C Systems

Most C2C systems evaluation lay emphasis on the effectiveness of the systems, i.e., whether the trust can be built and in which degree, moral hazard can be reduced. The most popular method for analyzing the effectiveness is game theory. *Game theory* concerns the actions of decision makers who are conscious that their actions affect each other (Rasmusen, 2001). It is a modeling frame which provides a set of tools that allow to analyze and predict how self-interested decision makers interact (Jurca, 2008).

A game is a situation in which players (participants) make strategic decisions that take into account each other's action and responses (Pindyck and Rubinfeld, 2009). As Rasmusen (2001) indicates the essential elements of a game are players, actions, payoffs, and information, which can be used to describe the situation by the modeler. Economists have extensively studied reputation in game theoretic settings (Mui et al., 2002a). Much of the economic studies on reputation relates to repeated games.

By using game theory in the online marketplace settings, scholars are able to analyze the effectiveness of reputation systems in C2C systems (Bolton et al., 2004; Whitmeyer, 2000) and the relationship between reputation information and product price (Jin and Kato, 2006; Houser and Wooders, 2006).

Zhou et al. (2009) used the data collected from eBay and analyzed the effec-

tiveness of a variety of feedback measures. eBay has at least four measures of their seller's feedback, i.e., the positive feedback counts, negative feedback counts, the overall feedback score (positive feedback counts minus negative feedback counts) and the percentage of positive counts. The authors argued that the positive and negative counts are ineffective at predicting the ending price in an auction.

Fan et al. (2005) showed that the simple accumulative and average score aggregation are not robust enough, sellers will lose incentives when their reputation scores are high enough and the transaction history is long enough. Pavlou and Dimoka (2006) concentrate on the role of the text feedback in C2C Systems. The authors argued that a single reputation rating is not sufficient for describing the reputation of a seller. However with the help of text feedback, more information can be obtained.

Gaur et al. (2010) assesses eight trust building reputation systems, including both commercial systems and academic models, centralized and distributed systems, based on their performances against different attacks. The authors defined 9 common problems and attacks against C2C Systems, such as, too many positive/negative ratings, low rated agents exists and re-enter the market. The results showed that most systems are only able to handle 3 or 4 attacks.

The primary job of C2C Systems is playing their sanctioning role to build trust between buyers and sellers, from this point of view, Resnick et al. (2000) then argued C2C systems should have the following properties:

- Entities long lived. One of the biggest problem for online marketplaces is that it is very easy for users to change their IDs. By doing that, the ID with a bad reputation can be easily abandoned and then changed to a new one. Friedman and Resnick (2001) suggested that by offering an entry fee or charge people for changing IDs may help to solve the problem. In contrast, Malaga (2001) argued that new comers may require some time to become familiar with a site, therefore, they should not be penalized for bad behavior as much as regular users.
- Feedback about current is captured and distributed (such information must be visible in the future).

- Past feedback supports users decision. Past feedback must have a direct reflection of one's past behaviour. However the authors also indicated that for auction websites, they may facing a problem that the buyers tend to negotiate with sellers before leaving negative feedback (Resnick et al., 2000).

Dellarocas (2001) argued that if binary feedback profiles are used to decide whether a seller advertises truthfully (in which case buyers assess quality equal to the advertised quality) or not, then, in theory, binary reputation systems can be well functioning. The author also claimed that the systems are expected to be quite fragile in practice except the system can supply more information to the buyers.

2.4.2 Review Centres

Research in the evaluation of review centres concentrates on the effectiveness of the product reviews, such as their influence on the sales and whether reviews can reflect the true quality of products.

Although product reviews may have a positive impact on the sales (which have been discussed in Section 2.3.2), whether they can reflect the true quality of products remains unclear. Review spam is the most concerning problem. There are two main kinds of review spams: untruthful reviews and non-reviews (Jindal and Liu, 2008; Hu et al., 2010). *Untruthful review*, a.k.a. fraud review, are the reviews that are posted intended to mislead readers. For example, an author or publisher may ask their friends to write extremely positive reviews for their books. *Non-reviews* literally means advertisement or any other irrelevant information that posted as reviews. Jindal and Liu (2008) argued that sometimes a review does not comment on the specific product but on the brand and it should be treated as spam as well; however the authors also admitted that these 'brand-only' reviews may be useful and provide extra information to others.

Several approaches can be adopted to eliminate or filter review spams. For example, a number of researchers have proposed some text mining and semantic analysis algorithms to reduce the number of non-reviews (Huang et al., 2010; Jindal and Liu, 2007; Lau et al., 2010).

The influence of untruthful reviews can be eliminated or at least hugely reduced by gathering reviews from sufficient sources (Fan et al., 2005; Resnick et al., 2000; Chen et al., 2004; Hu et al., 2006). For example, an author may find 10 friends to write positive reviews for one of his/her books, but when the book receives 1,000 reviews, their influences would be very small. Furthermore, some reputation systems also allow users to rate reviews as ‘helpful’ or not (Figure 2.7). Therefore, systems can sort the reviews based on the number of helpful votes.

In addition, some systems also allow users to help them filter the spam by reporting or flagging them (the ‘Report this’ in Figure 2.7).

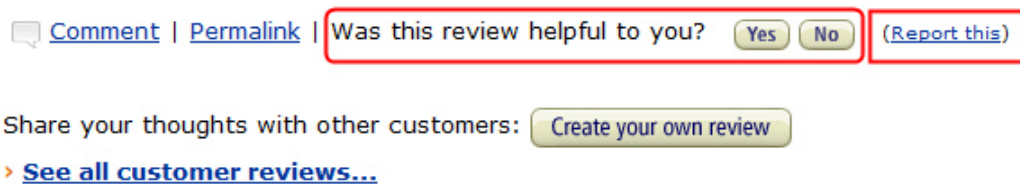


Figure 2.7: ‘Helpful’ Vote and Spam Report

Some researchers argue that even when there is no review spam, the self-selection bias will affect early reviews (Hu et al., 2008; Li and Hitt, 2008; Mudambi and Schuff, 2010). However as Li and Hitt (2008) indicated the problem can be solved by the larger number of reviews and the helpfulness votes from users.

In addition to the number of reviews, Davis and Khazanchi (2008) suggested several other characteristics that are important to online word-of-mouth systems, such as, the nature of the reviews, images uploaded by the reviewers and reviewer types. Li and Hitt (2010) indicated that Review Centres should collect multiple review dimensions to separate ‘perceived value and perceived quality’, because the authors found that the text reviews are more closely correlated to the perceived value rather than quality.

2.4.3 Online Communities

Online communities do not attract as much research as C2C systems or review centres.

Zhu (2010) analyzed the user behavior on Digg, and found that an overwhelming majority of users digg stories infrequently while a small number of users are very active. The author suggested that the majority may read an article first and then digg it if they like it, whereas the active users, who make a couple hundred diggs per day, may follow a different pattern. The paper also examined the social networking features of Digg. Digg like other social networkings allows users to designate others as friends, and the site provides a filtering mechanism which can highlight their friends activities. Zhu (2010) found that users with more friends tend to digg more and the friend interface influences users on viewing and rating contents. Lerman and Galstyan (2008) also focused on the social networking features on Digg, it suggested that the feature played a significant role in promoting stories.

Tran et al. (2009) pointed out that the effectiveness of a voting system, which used by online communities can be attacked by Sybil attack. Sybil attack refers to an entity that forges many identities and uses them to manipulate votes. The authors proposed an aggregation algorithm, which can resist the attack.

2.4.4 Cross-type Evaluation

A few researchers choose to compare different types of reputation systems. Liang and Shi (2005) focused on the rating algorithms. They first classified rating aggregation algorithms into five categories based on whether they weight ratings differently and how the weights are decided: average (AVG), half weighted (Half), weighted majority algorithm (WMA), personalized similarity measure (PSM) and Beta (the algorithm proposed by Jøsang and Ismail (2002)). Then the authors used a simulation tool to evaluate the algorithms performances, including algorithm complexity, system running cost, and system benefit. The results showed that most of the time, the better performance of system often come with higher system costs and complexity.

Ruohomaa et al. (2007) proposed a taxonomy including the creation and content of the information (rating and review), the selection and use of information source and the interpretation and reasoning applied to the gathered information. The paper assessed 11 systems, in which 10 of them are academic models. The authors

indicated that most reputation systems experiments tend to focus on prediction correctness instead of performance.

Sabater and Sierra (2005) provided a more general evaluation model. The authors selected several classification dimensions of trust and reputation systems including the conceptual model, information sources, information visibility types, granularity of the model, agent behavior, type of information and the reliability of the model. The paper then compared 13 academic trust and reputation systems, most of which are decentralized systems.

Hoffman et al. (2009) surveyed a number of academic models and commercial systems and measured their weaknesses to attack strategies. The authors developed a classification framework which divided reputation systems into three main processes: formulation, calculation and dissemination. *Formulation* measures the source of information and information type. The *calculation* assesses the aggregation algorithms, finally, *dissemination* considers the distribution and storage durability of information.

Malaga (2001) however focused on comparing different commercial sites. The paper reviewed 11 sites covered all three types of systems. With the analysis, the authors found some common problems of reputation systems: inaccurate algorithms, barrier to entry, no incentives to rate, inability to filter or search, no categorization and unlimited memory of information.

Most of these cross-type evaluation studies concentrated on decentralized systems and C2C systems. Little attention has been paid to review centres and online communities. Furthermore, much research has focused on academic models rather than commercial systems.

2.4.5 Successful Factors of Reputation Systems

Despite the lack of cross-type evaluation, three factors are commonly believed as the most influential ones of reputation systems (Fan et al., 2005; Resnick et al., 2000; Liu, 2006):

- Sufficient information sources.

Sufficient information sources is the key factor to assure reputation systems perform reasonably well (Resnick et al., 2000). A small number of ratings or reviews are easily attacked by review spams and then cannot reflect the true opinions from the sources (Buehgger et al., 2008; Davis and Khazanchi, 2008). Therefore, with a sufficient number of reviews, the overall rating can converge to the true quality (Chen et al., 2004; Liu, 2006).

- Those sources should provide unbiased information.

Theoretically, all the ratings and the reviews which are provided by information sources are biased, because they are subjective options. However, if a reputation system can collect sufficient information from different sources, the biases can be eliminated or at least hugely reduced (Fan et al., 2005; Buehgger et al., 2008). The next problem is the review spam, which has been discussed in Section 2.4.2. With the use of text mining/semantic analysis and the help from the users, reputation systems are expected to solve the problem. Huang et al. (2010) brought the problem back to the first factor, information source. The authors argued that the quality of the reviews are not only related to the content quality but also influenced by the credibility and granularity of the evaluators.

- The shared information has to be processed and presented in the most meaningful format.

Many studies, such as, Jøsang and Ismail (2002), Sabater and Sierra (2001), Aperjis and Johari (2010), Garcin et al. (2009) and Liang and Shi (2005), are carried out on the aggregation algorithms, which process ratings. However most of these algorithms are designed for C2C or distributed systems only. Moreover, as Liang and Shi (2005) indicated that a simple algorithm like the averaging aggregating was good enough. The authors also believed that the design of reputation systems should be emphasized on the dynamics of systems, rather than the rating aggregating algorithm.

Some researchers focused on the format of information. For example, Gregg (2009) compares the usefulness of numerical ratings of eBay and Amazon Mar-

ketplace, a C2C market with fixed prices. The author argues that eBay's binary rating scale ('-1', '0', '+1') seems less useful than Amazon's Likert-type scale (ratings from '1' to '5') when determining which sellers to buy from. However, the author also admits that the difference between the two may not be entirely due to the differences in the scale design as it could be influenced by sellers or products. Zheng and Jin (2009) argued that reputation systems should collect not only ratings and text reviews, but also multi-media information, such as pictures and videos.

Other researchers stated that numerical ratings cannot be sufficient to reflect the true quality of the products, while text reviews can be supportive by providing more detailed information (Li and Hitt, 2010; Pavlou and Dimoka, 2006; Ghose et al., 2005; Chevalier and Mayzlin, 2006). For example, Ghose et al. (2005) analyzed the text feedback in eBay and showed that they provide more information than numeric ratings. Furthermore, the paper pointed out that sellers may derive from different reputation dimensions, such as some sellers have a good reputation on fast delivery, while others may have good communicating skills. Chevalier and Mayzlin (2006) also showed that customers value the text reviews more than simple ratings. Li and Hitt (2010) suggest that single-dimension reputation systems are less effective than systems that collect multiple quality dimensions. In other words, one overall rating is not sufficient, and reputation systems should allow information sources to rate or review products from different quality dimensions.

With the consideration of the three common successful factors, there is still discussion on other relevant aspects. A number of researchers concentrated on the characteristics of information sources. For example, the credibility of the source is considered as an important factor. Malaga (2001) pointed out that to allow users to rate each other, not only can assess the credibility of them, but also can be seen as the incentives to the sources. In fact, Amazon has employed a similar mechanism, which allows users to vote product reviews as 'helpful' or not. Then the reviewers can be ranked by the percentage of the 'helpful' votes their reviews received. Huang et al. (2010) found that if a reviewer has a better credibility and expertise in the specific

area, their reviews may receive more helpful votes. Koh et al. (2010) analyzed the data collected from Chinese and American movie review websites and found that due to different cultures, the reviews from Chinese site have a better reflection on the true quality of the movies than the reviews from the US site.

Several researchers emphasized on time-related factors. For instance, Buchegger et al. (2008), Fan et al. (2005) and Malaga (2001) argued that reputation systems should be able to forget reputation over time or give different weights to aging information, which thus can emphasize the importance of behavior at one time over another. Dellarocas et al. (2004) found that some movie review sites give low weights to the reviews that are submitted within the first weeks of a movie's release in order to reduce the self-selection bias. Dellarocas (2006a) focused on the update frequencies of information and showed that under certain conditions, the cooperation and efficiency in C2C systems can be increased by reducing the frequency of users profile updates.

Other researchers pointed out that with the growing popularity of reputation systems, more and more ratings and reviews will be gathered. Therefore, reputation systems must consider an effective way to filter reviews for their users (Malaga, 2001; Huang et al., 2010; Zhang and Tran, 2011).

2.4.6 Summary

Based on the discussion in this section, Table 2.1 summaries the characteristics that are believed are important to reputation systems.

2.5 Evaluation of Information Systems

From a broader perspective, reputation systems are information systems (IS) that using the Internet as communication intermediary. Some researchers have attempted to use information system measurements to assess reputation systems. For example, Chen and Tseng (2010) adopts information quality (IQ) dimensions to evaluate the quality of reviews in reputation systems. They selected 9 IQ dimensions including believability, objectivity, reputation, relevancy, timeliness, completeness, appropri-

| Characteristics | | Related References |
|---------------------------------|--------------------|---|
| Source | Sufficient Sources | Fan et al. (2005); Resnick et al. (2000); Liu (2006); Buchegger et al. (2008); Davis and Khazanchi (2008); Chen et al. (2004) |
| | Credibility | Koh et al. (2010); Malaga (2001); Huang et al. (2010) |
| | Granularity | Huang et al. (2010); Malaga (2001); Sabater and Sierra (2005) |
| | Type | Koh et al. (2010) |
| Review Quality | | Jindal and Liu (2008); Hu et al. (2010); Huang et al. (2010); Jindal and Liu (2007); Lau et al. (2010) |
| Information Format | | Li and Hitt (2010); Pavlou and Dimoka (2006); Ghose et al. (2005); Chevalier and Mayzlin (2006) |
| Information Aggregation | | Jøsang and Ismail (2002); Sabater and Sierra (2001); Aperjis and Johari (2010); Garcin et al. (2009); Liang and Shi (2005) |
| Time-related | Timeliness | Buchegger et al. (2008); Fan et al. (2005); Malaga (2001); Dellarocas et al. (2004) |
| | Update Frequency | Dellarocas (2006a) |
| Information Filtering Mechanism | | Malaga (2001); Huang et al. (2010); Zhang and Tran (2011) |

Table 2.1: Successful Factors of Reputation System

ate amount of information, ease of understanding and concise representation. Then the authors proposed a model for assessing the quality of text reviews based on the dimensions.

Therefore, this section reviews literature in the IS evaluation area. The success of IS is multidimensional, includes information quality, system quality, service quality, use, user satisfaction and net benefits (DeLone and McLean, 1992; Seddon, 1997;

Delone and McLean, 2003). Figure 2.8, which reproduced from Delone and McLean (2003), illustrates the six measures and the interrelationship among them.

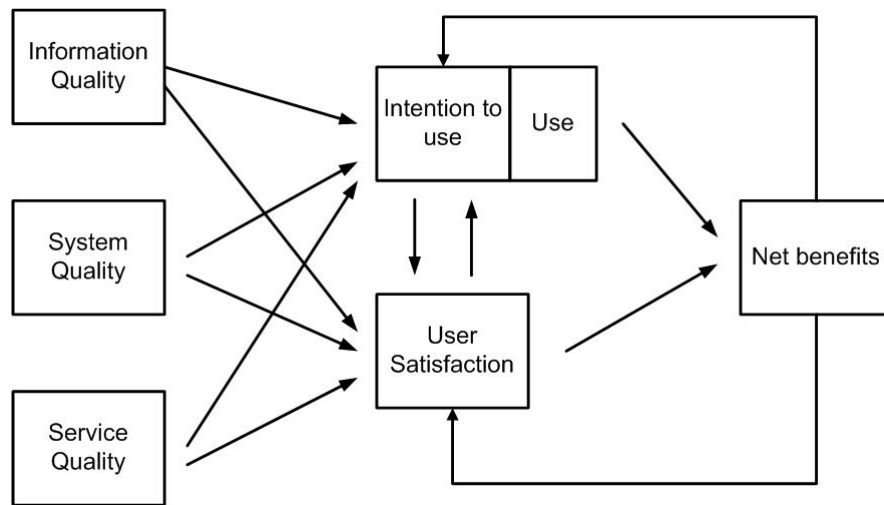


Figure 2.8: The D&M IS Success Model (Delone and McLean, 2003)

Information quality, system quality and service quality specify the technical qualities of IS. Use and user satisfaction interpret the success from system user perspective, whereas the net benefits reflects the effectiveness measures of IS. The arrows in the diagram showed how those measures influence each other. The basic technical qualities decide how users intend to use the system and their satisfaction. As a result the latter ones influence the IS impacts (net benefits).

Information quality measures the quality of the information and system quality concerns the characteristics of a system. Service quality (SVQ) measures the support the system users received from the IT department or the online company (Petter et al., 2008; Pather and Usabuwera, 2010), such as the effectiveness of online support and answers to frequently asked questions. However there is a debate on the measures of the SVQ (Petter et al., 2008; Pather and Usabuwera, 2010). Some researchers borrowed the idea from the marketing area and built an evaluation scale (SERVQUAL), including tangibility, reliability, assurance, responsiveness and empathy (Parasuraman et al., 2004; Pitt et al., 1995; Neill et al., 2001). Yang et al. (2003) and Tate and Evermann (2009), however, argued that SERVQUAL ‘does not provide a sound foundation for research into online service quality’ (Tate and Evermann, 2009, pg. 1).

This research only discusses the information quality and system quality because it focuses on the intrinsic nature of systems. Below is the review of the variables of each measure.

2.5.1 Information Quality (IQ)

The purpose of information systems is to gather, process and delivery information (Laudon and Laudon, 2007). Therefore the quality of information have a direct influence on the information systems success.

In general, information quality can be defined as ‘fitness for use’ (Wang and Strong, 1996; Kahn et al., 2002). However, like the evaluation of IS, IQ is also considered as multi-dimensional factor. Wang and Strong (1996) carried out an empirical study focusing on the consumers need of data quality. Based on the survey results, the authors developed a hierarchical model which presents the important aspects of information quality, including intrinsic, contextual, representational and accessibility. Redman (1997) identified 27 information quality dimensions and classified them into three groups: concept view, values and representation. Liu and Chi (2002) defined dimensions from the prospective of data evolution cycles: collection, organization, presentation and application qualities.

IQ literature provides a various and thorough classification of IQ dimensions; however there is no general agreement on the set of dimensions (Batini and Scannapieco, 2006). This research selected the five most important and widely accepted dimensions, which are relevant to the reputation systems, for further discussion: *accuracy, completeness, timeliness, accessibility and interpretability*.

2.5.1.1 Accuracy

In general, accuracy describes the degree of closeness of data content to its actual value (Wang and Strong, 1996; Chapman, 2005; Scannapieco and Catarci, 2002; Zhao et al., 2008). Some researchers specified that the accuracy is whether the data stored in the database (v) is conformance with real-world value (v') (Ballou and Pazer, 1985). However Redman (1997) argued that data accuracy should be defined from two aspects: the syntactic and semantic.

Under the Internet environment, it is believed that the reliability and credibility of information source are the key factors to assess the accuracy of information (Smith, 1997; Huang et al., 2010; Pernici and Scannapieco, 2002). As Katerattanakul and Siau (1999) and Metzger (2007) pointed out that with the advent of the Internet, the methodology of assessing Web information should consider the credibility of the information source as a property for measuring their ability to provide accurate information.

2.5.1.2 Completeness

Completeness is defined in the terms of the depth, breadth and scope of information (Wang and Strong, 1996; Liu and Chi, 2002; Pernici and Scannapieco, 2002). Bovee et al. (2003) specified that the completeness of the information should fit the information consumers' requirements. In other words, information consumers have different requirements, therefore, the acceptance degree of the completeness for different consumers are not always the same.

2.5.1.3 Timeliness

An important factor of information, particularly of the Internet information, is that it updates overtime (Tang et al., 2008; Batini and Scannapieco, 2006). Timeliness describes to which degree the information is up-to-date (Liu and Chi, 2002; Wang and Strong, 1996). For Web information, it means when the information has been submitted to the system. Some researchers may use different terms, such as, currency (Jarke et al., 1999; Redman, 1997) and age (Bovee et al., 2003), to express the same meaning as timeliness.

2.5.1.4 Accessibility

If information is not available to the information consumers, all the quality dimensions are irrelevant (Bovee et al., 2003; Zhao et al., 2008). Accessibility measures the ability of information consumers to access the information (Batini et al., 2009), i.e., whether the consumers are able to get the right information at the right time (Zhao et al., 2008; Scannapieco and Catarci, 2002). In addition, some researchers, e.g., No-

vak et al. (2000) and McKinney et al. (2002), argue that under online environments, accessibility also refers to the speed of access.

2.5.1.5 Interpretability

IS must present information in an easy-to-understand way with clear formatting (Wang and Strong, 1996; Scannapieco and Catarci, 2002). Bovee et al. (2003) pointed out that information consumers 'requirements for interpretability of information may be much broader'. For example, the ability for the users to customize the content is a vital factor for the e-commerce web sites (Delone and Mclean, 2004). It is important for these sites not only to present the information in a concise way but also to allow users to customize the given information.

2.5.2 System Quality (SQ)

From the traditional viewpoint, SQ identifies the desirable characteristics of IS, including ease of use, system reliability, system flexibility, functionality and system usefulness (Petter et al., 2008; DeLone and McLean, 1992).

However under the Web environment, the assessment of SQ should take the relevant Internet factors into consideration, such as security and connectivity. Therefore, in this section, the discussion will not only focus on the literature in the area of information systems but also in the e-commerce and website evaluation area. Below are the SQ dimensions which are suitable for online information systems.

2.5.2.1 Usability

Usability indicates the ease-of-use of the web sites, in other words, whether the users can browse and interact with the sites without difficulties (Palmer, 2002; McKinney et al., 2002; Spiller and Lohse, 1997). Henneman (1999) indicated that the usability of the web site need to fulfill the requirements of efficiency, effectiveness and user satisfaction. Yoon and Kim (2009) pointed out that a web site, in particular an online store must provide easy navigation, search and inquiry functions to users. Other researchers argued that websites should have concise, clear web design with simple and organized layout (McKinney et al., 2002).

2.5.2.2 Reliability

Reliability is one of the most important factors that influence the users satisfaction of the site (Liu and Arnett, 2000). It identifies the level of stability of the system (Straub and Carlson, 1989). Under the online environment, it refers to the rate of system failure and error occurrence (Yoon and Kim, 2009).

Sometimes, security is seen to be associated with the reliability (Aladwani and Palvia, 2002; Longstreet, 2010). Users usually worry about two main security problems: whether the system can safely keep their information (Yoon and Kim, 2009; Chen and Barnes, 2007) and whether the information sent by the system is secure (Koufaris and Hampton-Sosa, 2004). Security problems can hugely affect users trust of the system and influence their satisfaction (Molla and Licker, 2001).

2.5.2.3 Response Time

Research has found that even in the off line world, people have little patience (usually no more than 10 seconds) on waiting for systems response to their inquiries (Miller, 1968). With the development of Internet technologies and popularization of the broadband, the acceptable time for the loading time of a web page is considerably less (Hoover, 2006). Therefore, it is important for web sites to give quick responses to users' inquires.

2.5.2.4 Usefulness

Usefulness is another key dimension of SQ (Davis, 1989; Bailey and Pearson, 1983). Web users consider usefulness as a vital factor which influences their satisfaction (Yoon and Kim, 2009; Aladwani and Palvia, 2002). For e-commerce systems, usefulness specifies their transaction capabilities and customer feedback capability Delone and Mclean (2004); Palmer (2002).

2.5.3 Summary

Table 2.2 summaries the major IQ and SQ dimensions. IQ and SQ are the key measures of information system success. They represent the technical qualities of

| IQ Dimensions | SQ Dimensions |
|----------------------|----------------------|
| Accuracy | Usability |
| Completeness | Reliability |
| Timeliness | Response Time |
| Accessibility | Usefulness |
| Interpretability | |

Table 2.2: IQ and SQ Dimensions

information systems, which are the foundation of the user-perceived quality.

2.6 Conclusion

This chapter first surveys the literature related to the traditional notion of reputation. The concept of reputation has existed much longer than the Internet, however with the help of Internet technologies, reputation information can be stored and disseminated much more widely and for longer than in real world. From the emergence of the Internet, online reputation systems have been widely adopted by e-business sites and companies. However, there are several problems with current reputation systems analysis and evaluation research:

- There is a lack of cross-type evaluation. As discussed in Section 2.3.4 reputation systems can be classified into many different types. Most researchers focused on assessing and analyzing single type systems only. C2C Systems and Review Centres have attracted most attentions. Very few studies concentrated on proposing a comprehensive evaluation model which can comparing all kinds of systems.
- Within the few cross-type evaluation studies, most of them only concentrate on one aspect of reputation systems, such as aggregation algorithms. However, reputation systems not only aggregate ratings and reviews but also collect and disseminate them. Gregg (2009) has stated that even the same type of system may be different from one to another on the distribution and interpretation of

the feedback data. Therefore, a comprehensive framework which can evaluate and compare reputation systems under a set of representative and common conditions is in need (Sabater and Sierra, 2005).

- Much research in the area focused on proposing and analyzing academic models, which are suitable for decentralized systems only. Most studies that concentrated on commercial systems are conducted from social sciences perspective. Reputation system is a multi-disciplinary subject; however there are very few studies provide discussions and analysis from a computer science perspective (Sabater and Sierra, 2005). Analyzing reputation systems from the computer science perspective can provide a systematic and objective overview of the systems.
- It is surprisingly found with the literature that there is no systematic terminology has been proposed for reputation systems. Many researchers have pointed out that coherent classification and terminology is needed for reputation systems (Buechegger et al., 2008; Swamynathan et al., 2010).
- Another problem of current research is that the cost of reputation system has been long ignored. Many academic algorithms and models have been proposed but very few of them has considered the cost, such as system complexity and algorithm complexity.

Based on the discussion, this research will propose a comprehensive evaluation model, which focuses on the intrinsic nature of reputation system, for assessing different types of them. The model is built on the underlying structure of reputation systems.

Chapter 3

Reputation System Terminology and Structure

3.1 Introduction

This chapter analyses the intrinsic nature of reputation systems including essential entities and basic structure. Reputation systems may have different types, functions and interfaces, they all have the same underlying entities and structure (Friedman et al., 2007).

3.2 Essential Entities

Basically, reputation is the people's opinions about specific items.

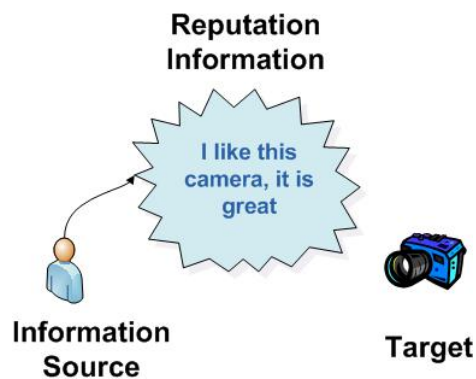


Figure 3.1: Essential Entities of Reputation System

Figure 3.1 shows that a person (information source) likes the camera (target). ‘I like this camera, it is great’ is the reputation information provided by the person.

Definition 1. (*Information Source*). *An information source provides information to a reputation system.*

Definition 2. (*Target*). *A target refers to the entity on which evaluators provide information.*

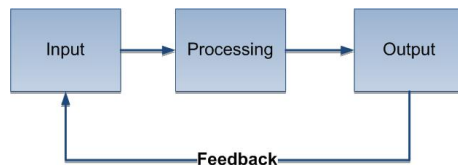
A target may be a product, a transaction or even a story.

Definition 3. (*Reputation Information*). *Reputation information refers to information related to a target’s reputation, such as reviews or ratings of a product.*

Essentially, reputation systems collect reputation information on targets from information sources, then aggregate and publish the information.

3.3 The Structure of Reputation Systems

In the area of Information Systems (IS), researchers tend to separate the system structure into four components (Figure 3.2a): Input, Processing, Output and Feedback (Laudon and Laudon, 2007). Input is the process of gathering data and Processing transforms raw data into information. Output then publishes information as meaningful output with certain formats, and Feedback is used to provide information to control the quality of the Input and Processing activities (Stair et al., 2010).



(a) IS model



(b) Model proposed from reputation system area

Figure 3.2: Different Structure Models

It is commonly accepted that reputation systems are a specific kind of information system. A number of researchers (Hoffman et al., 2009; Zheng and Jin,

2009; Swamynathan et al., 2010; Dellarocas, 2009; Friedman et al., 2007) proposed a similar structure for reputation systems (Figure 3.2b): Information Collection, Processing and Dissemination. Information Collection indicates the activities of collecting ratings and reviews from information sources, Processing refers to the aggregation of the ratings, and Dissemination refers to the distribution of reputation information.

Comparing the two models, it can be found that the reputation system model (Figure 3.2b) has ignored the Feedback, which is proposed by IS model. Feedback is a vital component that can control the quality of the Input and Processing by providing feedback information on the Output. Reputation systems can use the feedback to assess the quality of ratings and reviews. For example, Amazon allows their users to vote on product reviews as ‘helpful to you’ or ‘not helpful’. These reviews can then be ranked by the number of ‘helpful’ votes they have received.

In addition, the two models have both ignored another important component — the storage. As discussed in Section 2.2, impermanency is a big obstruction of off line reputation dissemination, whereas online reputation systems can store information for long term uses. In other words, the storage of reputation information must be considered as a basic component of reputation systems.

Therefore, based on the above analysis, regardless of their interfaces, functions and types, all reputation systems should have five underlying components: Input, Processing, Output, Feedback Loop and Storage.

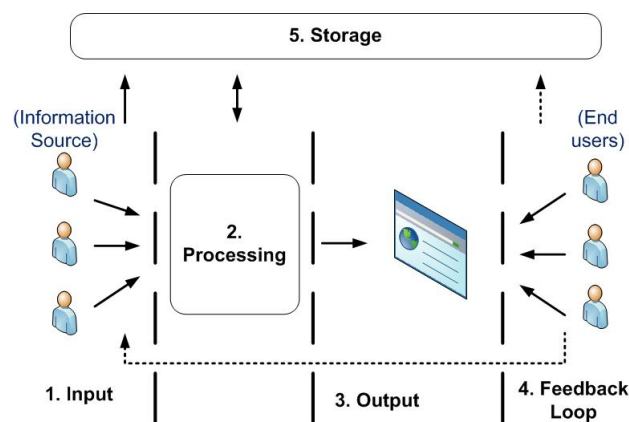


Figure 3.3: Reputation System Structure Model

Figure 3.3 presents the interrelationships between the five components. It shows that reputation information flows from sources to the processing component. After being aggregated, it will be published. If end users are interested, they may be allowed to leave feedback (the dotted lines indicate that the feedback loop is an optional component). During the whole process, all information needs to be stored.

3.3.1 Input

Input is the activities of collecting reputation information about the targets and other related information from information sources.

It should be noted that most of the time information is provided by evaluators (Figure 3.4).

Definition 4. (*Evaluator*). *An evaluator is a person who provides reputation information.*

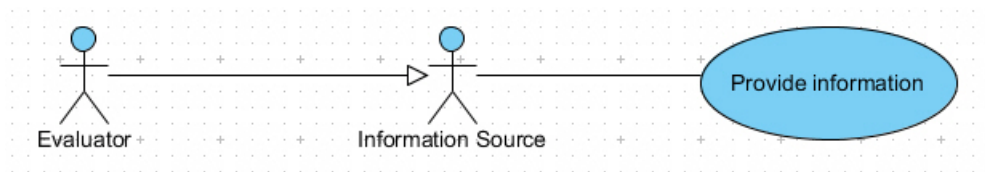


Figure 3.4: The Use Case Diagram of Evaluator and Information Source

In some cases, reputation systems can collect information from other reputation systems rather than from evaluators directly.

Reputation systems usually collect two main kinds of reputation information: explicit information and implicit information.

Definition 5. (*Explicit Information*). *Explicit information indicates the information that evaluators actively provide, such as a rating on the product or a text review.*

Definition 6. (*Implicit Information*). *Implicit information is usually generated from evaluators' activities. For example, the total number of views of a video or a book sales figure.*

The problem with the implicit information is that the true opinions of all the evaluators may not be reflected in the information. For example, a person may

buy a book which eventually they dislike. Furthermore, when buying the book they may not be aware of the consequences of their activities, while evaluators, who provide ratings and reviews, know that the information will have an influence on the target's reputation. Usually implicit information has more 'evaluators' than explicit information, because not all evaluators provide explicit reputation information. For instance, not all the people who buy a product will provide a rating. This research concentrates on explicit information only.

In addition to information on the target, reputation systems also collect other related information, such as, information about the evaluators. Gathering more information about the evaluators can help end users build trust on the provided information. In particular, it would be an advantage if a system can identify the credibility of evaluators. The credibility of evaluators can be specified by many means and one of the most popular approaches estimates it based on the quality and quantity of the reputation information they have provided. Credibility information is collected from *evaluator credibility providers* (Figure 3.5).

Definition 7. (*Evaluator Credibility Provider (EC Provider)*). *The information collected from an EC Provider can be used to generate the credibility level of the evaluator.*

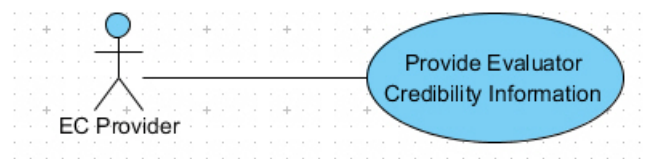


Figure 3.5: The Use Case Diagram of EC Provider

There will be more discussion on EC Providers in the Section 4.2.2.

3.3.2 Processing

Processing is the procedure of computing and aggregating the reputation information.

After collecting information, reputation systems need to aggregate it into a meaningful form. For example, numeric ratings can be summed or averaged to an overall

rating which can represent major evaluators' opinion on the target. Processing can also provide other functions of the whole system.

3.3.3 Output

Output indicates the dissemination of the reputation information.

After collecting and aggregating the information, reputation systems need to present the information to the end users (Figure 3.6).

Definition 8. (*End User*). *End users are the people who use the reputation systems for seeking information about a target.*

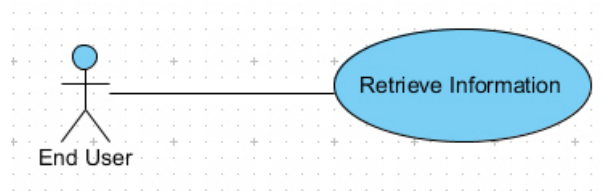


Figure 3.6: The Use Case Diagram of End User

The aim of reputation system is to provide information for end users, so that they can make decisions based on the collected and processed information.

3.3.4 Feedback Loop

A feedback loop is the collection of the feedback about the review, which can be seen as the 'review of the review'.

The quality of reviews is an important factor to reputation systems. Reviews are provided by many evaluators and it is difficult to measure the quality of reviews before they are published to the website. As discussed earlier, reputation systems usually choose to use feedback ('helpful' votes on the reviews) to control the quality of reviews. The users who provide feedback are the feedback providers.

Definition 9. (*Feedback Provider*). *A feedback provider is the person who leave feedback on the reviews.*

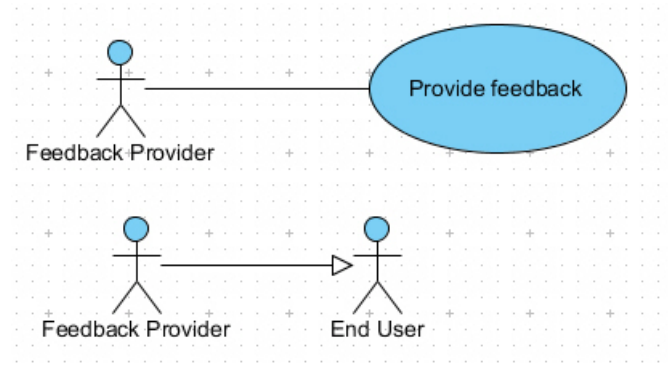


Figure 3.7: The Use Case Diagram of Feedback Provider

Figure 3.7 is the use case diagram of a feedback provider. Feedback providers are supposed to give ratings on reviews based on their quality, which means they are able to retrieve the reviews. In other words, feedback providers are end users.

Reputation systems do not always have feedback loops, i.e., it is an optional component. It should be noted that some websites use the word ‘feedback’ to refer to the reviews (reputation information). To avoid confusion in this thesis, the word ‘feedback’ is used to indicate the information collected in the Feedback Loop component only.

3.3.5 Storage

The storage refers to the process of storing all the collected and processed information.

Information that is collected, processed and published within the system can all be stored. The storage enables reputation information to be retrievable for a very longer time.

3.4 Conclusion

This chapter analyses essential reputation systems entities, including basic users and underlying system structure components. The main users of reputation systems are: evaluators, end users, EC providers and feedback providers. Figure 3.8 summaries the roles and relationships of these users. EC and feedback providers need to ac-

cess the reputation information (i.e., read reviews) before they can provide relative information. In other words, they are also end users.

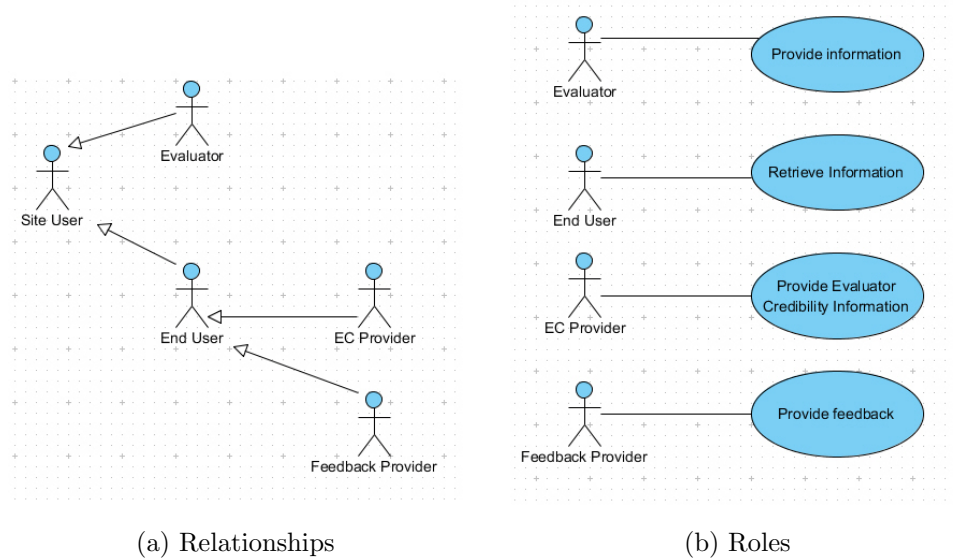


Figure 3.8: The Use Case Diagram of System Users

All reputation systems, no matter their different interfaces and functions, can be divided into five components: input, processing, output, feedback loop and storage. Figure 3.9 illustrates the interrelationship among the entities and components of the reputation system. It shows that the information collected from the sources are aggregated by the processing component and then published to the end users.

Based on the different requirements and characteristics of each component, a series of benchmark criteria can be defined. Therefore, reputation systems can be assessed regardless of their different interfaces or functions.

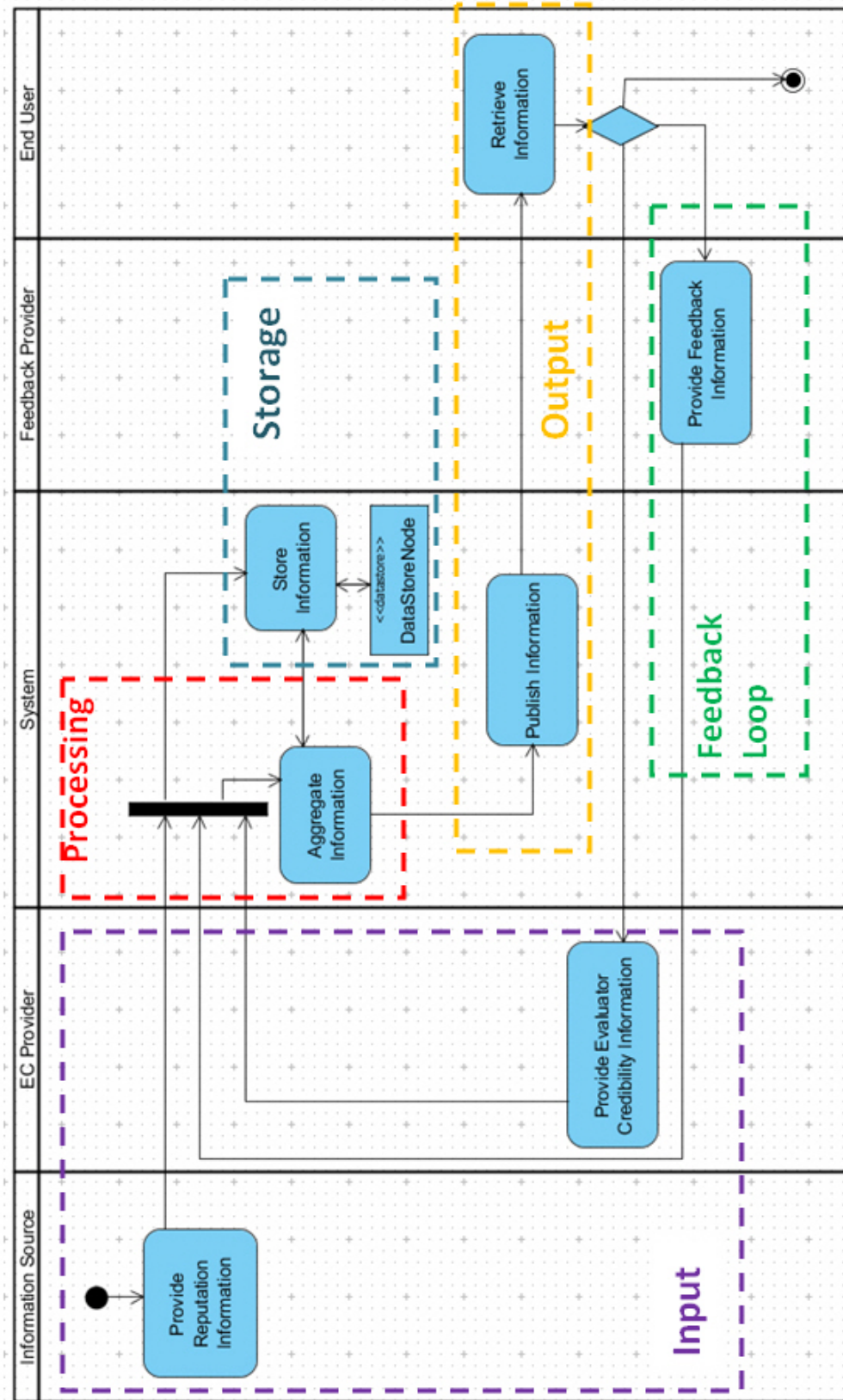


Figure 3.9: The Activity Diagram of Reputation System

Chapter 4

The SERS Model

4.1 Introduction

This chapter introduces the SERS (Systematic Evaluation of Reputation System) model, which aims at systematically evaluating different types of reputation system. As discussed in Section 3.3, reputation systems have five underlying components: input, processing, output, feedback loop and storage. The SERS is built on these components. It defines a number of criteria according to each component's characteristics.

Before further discussion, it should be noted that online reputation systems do not solely exist on the Internet. They are integrated within commercial websites. Thus some factors of the websites, such as web page design and usability, may also have influences on the performance of reputation systems. As this research focuses on the intrinsic nature of reputation systems, the factors of the websites will not be considered.

Section 4.2, 4.3, 4.4, 4.5 and 4.6 introduces the criteria defined on each component respectively. They first analyze the nature of each component and their specific characteristics, then discuss the relevant criteria including their definition and possible measurements.

4.2 Input

Input refers to the collections of ratings, text reviews and other relevant reputation information. It is one of the most important components because the other components rely on the information collected from it.

Three essential elements are involved in the input: information source, collection channel and reputation information. Reputation systems use *collection channels* to gather *reputation information* from *information sources*.

4.2.1 Collection Channel

Criterion I 1. *Collection Channel*

Collection channels refer to the approaches of collecting information from the sources, i.e., how the systems collect information. There are two main kinds of channels: direct channels and indirect channels. *Direct channels* collect information directly from the evaluators. Within these, some systems passively wait for the evaluators to write reviews (Channel C_{1a}). Other systems choose to invite evaluators via email or web page links (Channel C_{1b}). Figure 4.1 is the UML sequence diagram for the direct channels.

Unlike direct channels, *indirect channels* collect reviews from other reputation systems (Channel C_2). For instance, A number of reputation systems have agreed to allow Google to retrieve their reviews and publish partial or full reviews on the Google Shopping page.

It should be noted that it is possible for a system to use a combination of collection channels.

4.2.2 Information Sources

Information sources are important to reputation systems because they provide information. The nature of the source is an important factor when assessing the reliability and credibility of information (Katerattanakul and Siau, 1999; Wang and Strong, 1996).

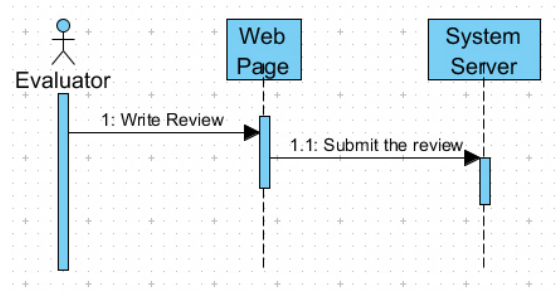
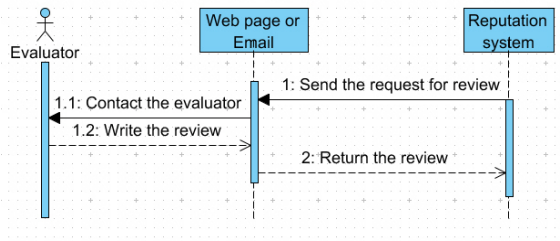
(a) Channel C_{1a} (b) Channel C_{1b}

Figure 4.1: The Sequence Diagram of Direct Channels

Criterion I 2. *The Set of Evaluators*

It is essential for a target to get a sufficient number of ratings/reviews before reputation can reflect its true quality (Dellarocas, 2003; Resnick et al., 2000). Each target can attract a set of different evaluators (U_e). The size of U_e can be calculated by:

$$|U_e| = |U_q| * p_e \quad (4.1)$$

U_q is the set of people who are qualified to leave reputation information, i.e., those who are eligible to be evaluators. Not all the evaluators will leave ratings or reviews. Considering that when sending out surveys, only a small number of which will be returned. Only a small number of eligible evaluators will actually leave reviews as well. Thus, p_e denotes the proportion of people who actually provide reviews, which is similar to the response rate in surveys.

The number of eligible evaluators ($|U_q|$)

Who are eligible to be evaluators depend on the systems' regulations. For example, Amazon allows evaluators to leave reviews on any products, once they have

registered with the site and bought one item. However eBay only allows the parties of the transaction (buyers and sellers) to rate each others. U_q can be classified into five sets (Figure 4.2).

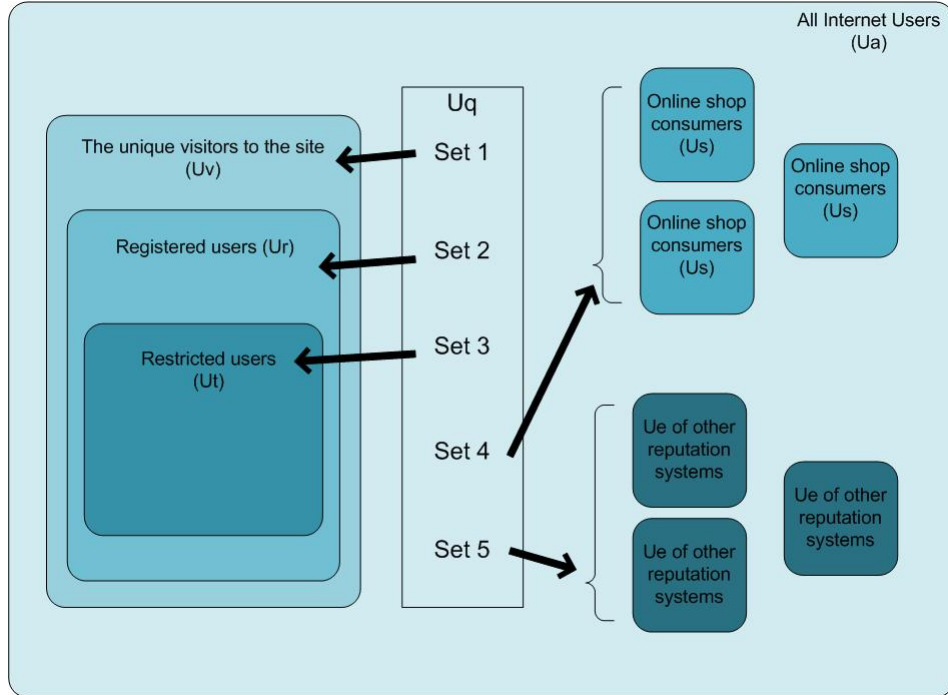


Figure 4.2: The Sets of U_q

First, consider systems that limit evaluators to their own site visitors. The first choice a systems has is to allow all their visitors (U_v) to be evaluators, which means, $U_q = U_v$ (*Set1*). A more common case is that the system requires evaluators to register first, i.e., the system accepts all registered users (U_r) to be evaluators, $U_q = U_r$ (*Set2*). Moreover, systems may limit their evaluators to a smaller set (U_t) with further restrictions. For example, Reevoo, a product review center, asks evaluators to provide a proof of purchase before leaving product reviews. Thus, $U_q = U_t$ (*Set3*).

Second, in addition to collecting reputation information from own site users, systems may collect reviews from the users of other sites. Some reputation systems work with a number of online shops, which allow the system to collect reputation information from their customers after purchases. Therefore, the eligible evaluators are the summation of all the shop's customers: $U_q = \sum_{i=1}^{N_s} U_{s,i}$ (*Set4*). $U_{s,i}$ is the set of customers of the i th shop and N_s is the number of shops that have cooperated

with the reputation system.

As discussed in Section 4.2.1 some systems, such as Google Shopping, use collection channel C_2 to collect information from other reputation systems rather than from evaluators. In other words, this kind of system retrieves and combines cooperated systems' information on their own platforms (*Set5*). Therefore, $U_q = \sum_{i=1}^{N_r} U_{e,i}$. N_r is the number of cooperating reputation systems and $U_{e,i}$ denotes the number of evaluators of the i th cooperated system.

In summary:

$$|U_e| = \begin{cases} |U_v| * p_e & \text{Set 1: all system visitors can be evaluators} \\ |U_r| * p_e & \text{Set 2: if only registered users can be evaluators} \\ |U_t| * p_e & \text{Set 3: if only people have registered and are} \\ & \text{qualified for further restrictions can leave reviews} \\ \sum_{i=1}^{N_s} (|U_{s,i}| * p_{e,i}) & \text{Set 4: if systems cooperate with online shops} \\ \sum_{i=1}^{N_r} |U_{e,i}| & \text{Set 5: if systems collect information} \\ & \text{from other reputation systems} \end{cases} \quad (4.2)$$

The proportion of people who actually leave reviews (p_e)

The proportion of people who actually leave reviews (p_e) can be influenced by many factors. For example, the collection channel is considered to have an influence on the p_e , which is because the evaluators are more likely to leave reviews if they receive a reminder from the system (systems use C_{1b} channel). Web page design, the nature of target and even the nature of the website can also have impacts on p_e .

The number of reviews a target can receive (N_{tr})

The size of the set of evaluators ($|U_e|$) can be used to estimate the number of reviews that the target can receive (N_{tr}). Most of the time, reputation systems only allow an evaluator leave reputation information to the same target once, thus, $N_{tr} = |U_e|$. However, some systems, in particular C2C systems, allow buyers and sellers

leave reviews about each other after every transaction. That is to say, evaluators can leave reputation information on the same target repeatedly. It can be imagined that one may take advantage of this policy to increase own ratings rapidly, by exchanging ratings after fake transactions with the same person. To avoid this problem, reputation systems usually do not count every rating for the overall rating score. For example, in eBay if a seller receives multiple ratings from the same buyer within the same week, the seller's reputation score will only be affected by 1 rating¹. Therefore, for these systems, $N_{tr} = C * |U_e|$. C is the factor to calculate multiple ratings.

In general the set of evaluators have a great influence on the number of reviews a target can receive, which is the key factor in deciding whether the reputation information can reflect the target's true reputation.

Criterion I 3. *Granularity*

Granularity identifies how evaluators associate with targets. There are two kinds of granularities between an evaluator and a target.

- *The expertise granularity* refers to the evaluator's level of expertise in the target's area. An individual may enjoy a high reputation for their expertise in one domain while having a low reputation in another (Zacharia and Maes, 2000). For reputation systems, the expertise of an evaluator can be illustrated by their credibility in the same domain. For example, if an evaluator has good credibility for writing reviews on digital cameras, they are supposed to have a high level of expertise granularity with digital cameras.
- *The interaction granularity* indicates whether the evaluator has any direct interactions with the target. If a person has interaction with the target, for example, owns the product, their opinions are more believable than the others.

A good reputation system should be able to identify the granularity between evaluators and the targets. Thus this criterion indicates in which degree the reputation

¹eBay (2011), 'eBay help page'. <http://pages.ebay.com/help>; Last Accessed 15 January 2011.

system can identify the granularity.

1. A system has a *low level of granularity* if it cannot specify the evaluator's expertise credibility in specific area nor is it able to identify the interaction granularity.
2. A system has an *interaction level of granularity* if it is able to identify the interaction granularity between the evaluators and the targets.
3. A system has an *expertise level of granularity* if it can specify the evaluators expertise credibility.
4. A system has a *high level of granularity* if it can not only specify the expertise granularity but also identify the interaction granularity.

To specify the expertise granularity, reputation systems can use the results of the feedback loop to calculate the credibility of evaluators in different categories. If a system can present whether an evaluator has interaction with the target, for example, whether the evaluator has owned the product, it then can identify the interaction granularity.

Criterion I 4. *Evaluator Credibility*

It is important for reputation systems to have the ability of assessing the evaluators credibilities (EC), which can be seen as the reputation of the evaluator. In reputation systems, an evaluator's credibility is associated with the quality and quantity of the reviews they have written.

This section concentrates on the EC providers and Criterion *P2* in Section 4.3.1 identifies how the credibilities are calculated. Within the entities that have been discussed in Chapter 3, three of them can be EC providers: feedback providers, targets and the end users.

1. Feedback Providers. Some systems allow end users to give feedback on the reviews. The results of the feedback influence the credibility of evaluators. For example, Amazon lets end users rate the reviews as 'helpful' or 'not helpful'. The evaluator's credibility score will rise with the increase of the 'helpful' votes they received.

2. Targets. In C2C systems, where buyers rate sellers, the sellers also have opportunities to rate the buyers, which means that the rating given by the seller (the *target*) influences the credibility of the buyer (the *evaluator*). Then each agent's score can be seen as a reputation score or a credibility score.
3. Reputation systems can also allow end users to rate the evaluators on their credibility directly. For example, end users may rate an evaluator as a trustable evaluator, then the higher score the evaluator gets, the better credibility they have.

It should be noted that some systems have a special ranking mechanism for their users, called the 'Karma' mechanism. It records every activity a user has done within the system, then gives points to it (Farmer and Glass, 2010). For example, Yahoo! Answers is a website which allows people to ask and answer questions within the community. Each time users answer a question they will get 2 points. With this Karma mechanism, users have scores. Usually, the higher the score, the more active they are in the community. Because most sites use Karma mechanisms to identify the behavior of evaluators rather than reflect the credibilities of evaluators; therefore, the research does not consider it as a credibility mechanism.

4.2.3 Reputation information

Criterion I 5. *Information Format*

When collecting information from evaluators, reputation systems usually supply a form for evaluators to fill in (like a survey, see Figure 4.3). It contains information with different formats, including ratings, text comments or even rich media (photos and videos) formats. This criterion specifies the different information formats a reputation system accepts.

Information can be collected, presented and stored with different formats. The main formats are: numeric ratings, text reviews and rich media formats. Different information formats have different roles in reputation systems. For example, numeric ratings can be aggregated, so that when presenting it to end users, it can provide a comparable meaning. While text reviews contain more detailed information. Rich

* Rate this hotel: Very Good

Cleanliness Location Rooms

Service Value Sleep Quality

* Write your review

* Date of stay
Select one

* Purpose of trip
 Business Leisure

* Traveled with
Select one

* Title your review – If you could say it in one sentence, what would you say?
“ ”

* Your review [Tips & guidelines](#) (50 character minimum)

* Would you recommend this hotel to a friend?
 Yes No

Figure 4.3: Snapshot of Tripadvisor ‘Write a Review’ page

media formats refer to the pictures or videos, which can give a better illustration of the target.

Criterion I 6. *Information Breadth*

Information breadth identifies the number of properties that have been collected. The breadth is an important dimension for assessing the completeness of information. More information can illustrate a clearer image of the target. For example, Tripadvisor.com, a travel-related review center, encourages their evaluators to rate hotels for their ‘value’, ‘rooms’, ‘location’, ‘cleanness’ ‘sleep quality’ and ‘service’ separately.

Although end users may desire more information, too much information may reduce the evaluators’ motivation for completing reviews. Reputation systems can let evaluators choose how much information they want to provide by marking the properties as ‘*Required*’ and ‘*Optional*’.

Take Tripadvisor as an example. Figure 4.3 shows that the site requires evaluators to provide 7 pieces of required information (the required information is labeled with a red star), including 5 ratings and 2 text reviews. It also requests 6 optional

ratings. It means that an evaluator can choose to provide all 13 pieces of information or just the 7 required ones.

4.2.4 Collection Costs

Criterion I 7. *Input Collection Costs* (T_{ip})

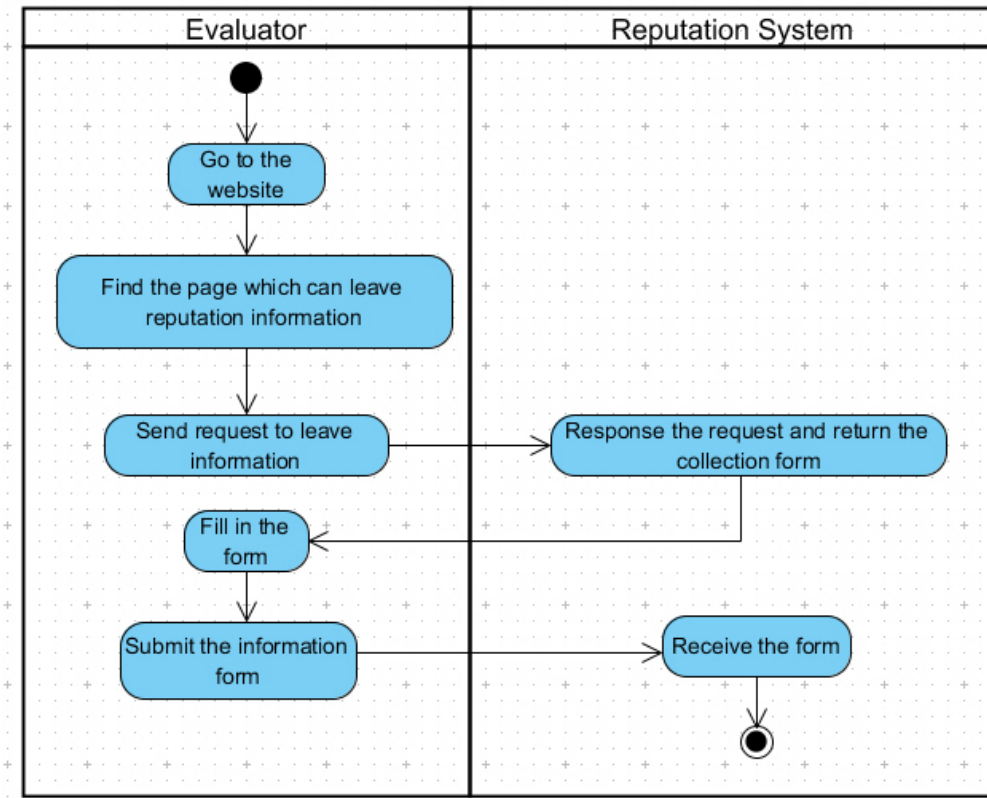
As noted earlier, the aim of this research is to evaluate reputation systems rather than the performance of the website, which means the costs of building or promoting a website is of no concern. Thus this criterion refers to how much time it takes to collect a single unit of reputation information. The collection channel decides how the information is collected; therefore, it causes different cost. For C_{1a} and C_{1b} systems, reputation information is provided by evaluators, which means, the collection cost is how much time it takes an evaluator to complete the ‘collection form’. C_2 systems collect information from other systems, the cost is therefore depended on the indexing speed of the system.

4.2.4.1 Collection Costs for C_{1a} Systems

In order to calculate the collection cost of C_{1a} systems, it needs the process that the evaluator has to use to leave reputation information. The evaluators of C_{1a} systems actively provide reputation information by browsing the system’s web pages. When they decide to give ratings or reviews, evaluators need to browse the web site to find the page where they can leave their reviews and then provide the information.

Figure 4.4 is the activity diagram of the process of an evaluator provide reputation information. It shows that an evaluator first browses several pages until they find the pages that they can provide information. Then the system shows the collection form for them on request. The evaluator then fills in the form and submits it to the system. Therefore, the collection cost for C_{1a} systems ($T_{ip,c_{1a}}$) is the time it takes for the whole processes.

$T_{ip,c_{1a}}$ can be calculated by the time an evaluator needs to browse the web site (T_{br}) plus the time it takes them to fill in the collection form and submit it to the system (T_{cp}).

Figure 4.4: The Activity Diagram of C_{1a} System

$$T_{ip,c_{1a}} = T_{br} + T_{cp} \quad (4.3)$$

The estimation of T_{br}

The browse time can be estimated by the page loading time and the time the evaluator needs to read the content:

$$T_{br} = (T_{ld} + T_{rd}) * N_{pg} \quad (4.4)$$

T_{ld} is the time for loading one web page, T_{rd} denotes the time the evaluator needs to browse one web page and N_{pg} is the total number of pages that the evaluator needs to browse.

As T_{br} is related to business strategies rather than the reputation system itself, the following assumptions can be made to simplify the formula:

- With the development of internet technologies, T_{ld} is a very small number

when comparing to human reading/inputting time. Thus it can be assumed:

$$T_{ld} = 0.$$

- The browsing time is related to the number of words on the page and the reading speed of the evaluator. Weinreich et al. (2008) has proposed a formula to calculate the average browsing time according to the number of words (W_{pp}) on a page:

$$T_{rd} = 0.044 * W_{pp} + 25.0 \quad (4.5)$$

Assume all systems have the same $W_{pp} = 200$, then:

$$T_{rd} = 0.044 * 200 + 25 = 33.8(\text{seconds}) \quad (4.6)$$

- Suppose all systems require users to browse two pages: $N_{pg} = 2$.

According to Equation (4.4–4.6),

$$T_{br} = (0 + 33.8) * 2 = 67.6 \quad (4.7)$$

The estimation of T_{cp}

The time for the evaluator to complete the collection form depends on the format and the amount of information,

$$T_{cp} = \sum_{j=1}^{N_{if1}} T_{cp,if1,j} + \sum_{j=1}^{N_{if2}} T_{cp,if2,j} + \sum_{j=1}^{N_{if3}} T_{cp,if3,j} \quad (4.8)$$

N_{if1} , N_{if2} and N_{if3} are the number of ratings, text comments and rich media format information respectively. $T_{cp,if1}$, $T_{cp,if2}$ and $T_{cp,if3}$ denote the time for completing the corresponding information.

- $T_{cp,if1}$: To rate a target, a mouse can be used to make the selection. According to Hansen et al. (2003), the time for completing a task by mouse is between 0.93s – 1.45s, on average, it is 1.2 seconds.

- $T_{cp,if2}$: The time to write a text review depends on the words to be written (W_{pr}) and the human input speed. For general computer users, the average rate for composition is 19 words per minute (Karat et al., 1999). Then the $T_{cp,if2}$ can be calculated by:

$$T_{cp,if2} = W_{pr} * \frac{60}{19} \quad (4.9)$$

- $T_{cp,if3}$ is the time for creating and uploading a picture or video, which depends on the size of the file and the speed of internet connections. It is considered to be much larger than $T_{cp,if1}$ and $T_{cp,if2}$. This cost could vary according to different conditions, thus further assumptions and estimations would be inappropriate.

Then based on the above analysis, Equation (4.8) becomes:

$$\begin{aligned} T_{cp} &= \sum_{j=1}^{N_{if1}} 1.2 + \sum_{j=1}^{N_{if2}} 3.16 * W_{pr,j} + \sum_{j=1}^{N_{if3}} T_{cp,if3,j} \\ &= 1.2 * N_{if1} + 3.16 * \sum_{j=1}^{N_{if2}} W_{pr,j} + \sum_{j=1}^{N_{if3}} T_{cp,if3,j} \end{aligned} \quad (4.10)$$

Therefore, according to (4.4)–(4.10), $T_{ip,c1a}$:

$$\begin{aligned} T_{ip,c1a} &= T_{br} + T_{cp} \\ &= 67.6 + 1.2 * N_{if1} + 3.16 * \sum_{j=1}^{N_{if2}} W_{pr,j} + \sum_{j=1}^{N_{if3}} T_{cp,if3,j} \end{aligned} \quad (4.11)$$

Practically creating a rich media review takes much more time than generating a text review or making a rating. Thus according to Equation (4.11), the domination factor is the $T_{cp,if3,j}$. However at the moment, most systems do not accept rich media information and as the cost is much higher, very few evaluators tend to provide it either. In this case, $W_{pr,j}$ becomes more decisive.

4.2.4.2 Collection Costs of C_{1b} Systems

Systems using collection channel C_{1b} send invitations to evaluators rather than passively waiting for evaluators to leave reviews. Within the invitation, systems usually provide with the link of the review page directly.

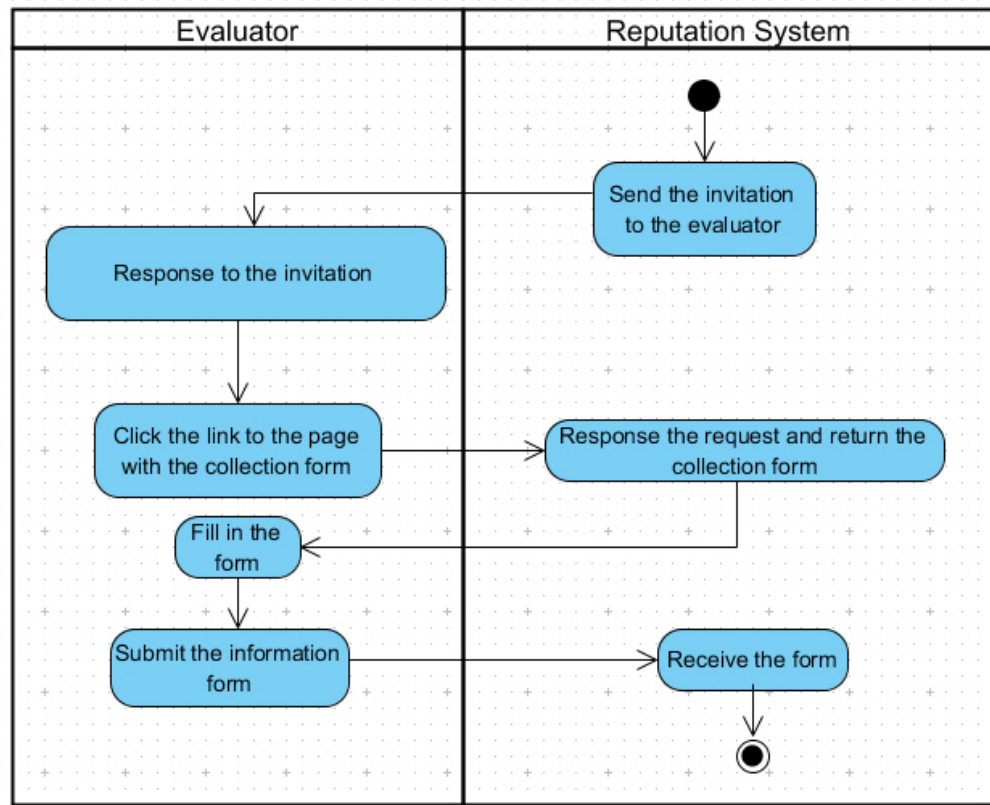


Figure 4.5: The Activity Diagram of C_{1b} System

Figure 4.5 illustrates the process for C_{1b} evaluators to provide information. It shows that systems first send invitations, such as emails or web links, to evaluators, then wait for the evaluators to respond. It may take evaluators minutes, hours, days or even weeks to respond to the invitation. When evaluators decide to provide information, they then go to the web page and fill in the collection form.

Similarly to the estimation of C_{1a} systems, the collection cost of C_{1b} systems ($T_{ip,c_{1b}}$) is calculated from the moment the evaluator decides to provide information, i.e., the moment they respond to the invitation. Thus, the collection costs for C_{1b} is:

$$T_{ip,c_{1b}} = T_{cp} \quad (4.12)$$

According to Equation (4.8)–(4.10):

$$T_{ip,c_{1b}} = 1.2 * N_{if1} + 3.16 * \sum_{j=1}^{N_{if2}} W_{pr,j} + \sum_{j=1}^{N_{if3}} T_{cp,if3,j} \quad (4.13)$$

4.2.4.3 Collection Costs of C_2 Systems

Collection channel C_2 collects information from other reputation systems (Figure 4.6), which means, the collection is done by machines.

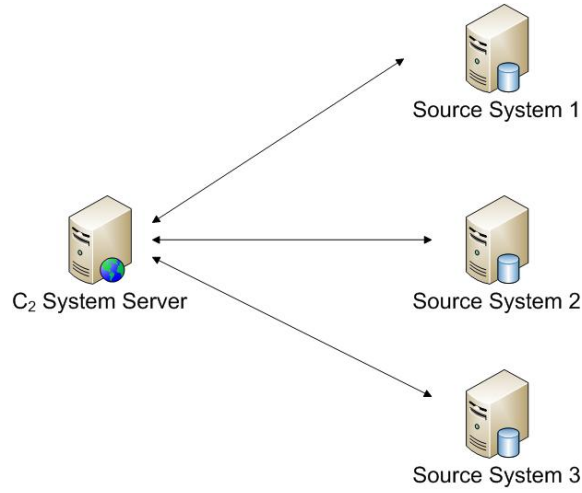


Figure 4.6: C_2 System Collection Process

Therefore, collection costs can be estimated as their information tracking time only:

$$T_{ip,c_2} = \text{indexing speed} \quad (4.14)$$

Usually it only takes no more than one second to get one piece of information.

4.2.4.4 Summary

The equations showed that for C_{1a} and C_{1b} systems, the information format (Criterion $I5$) and breadth (Criterion $I6$) have a big impact on the costs. Under the

current technology conditions, it takes much more time to collect a rich media format information than ratings or reviews. However the cost for C_2 systems, which is much lower than the other channels', mainly depends on the performance of their indexing technologies.

4.3 Processing

Processing is a set of activities that transforms raw information into more meaningful forms. The first three criteria, target rating algorithms, evaluator credibility (EC) algorithms and feedback aggregation algorithms, specify the algorithms that reputation systems used to compute the corresponding information. The other criteria, update frequency, robustness, algorithm complexity and system complexity, identify the attributes of these algorithms.

4.3.1 Algorithms

Reputation systems usually collect numeric ratings from evaluators, EC providers and feedback providers, then aggregate these ratings to one or several overall ratings to illustrate the opinions of the majority.

Criterion P 1. *Target Rating Algorithms*

This criterion specifies which algorithms are used by a system for aggregating the overall ratings of the targets. Assume an evaluator provides a rating r_i , then the overall rating is:

$$O(r) = f\{r_1, r_2, \dots, r_n\} \quad (4.15)$$

Criterion P 2. *EC Algorithms*

Reputation systems usually collect EC information as ratings as well. This criterion indicates how the EC information is aggregated. The overall EC rating is generated from the individual ratings (c_i):

$$O(c) = f\{c_1, c_2, \dots, c_n\} \quad (4.16)$$

Criterion P 3. *Feedback Aggregation Algorithms*

This criterion identifies how feedback loop ratings are aggregated. Again, the overall feedback rating is:

$$O(d) = f\{d_1, d_2, \dots, d_n\} \quad (4.17)$$

4.3.1.1 Common Algorithms

There are many ways to calculate these ratings. In statistics, a measure of central tendency, such as algorithmic mean, mode or median, is usually a value that best describes some attribute of the population. Some researchers have proposed more complex aggregating algorithms, for instance, Bayesian systems (Whitby et al., 2004; Jøsang and Ismail, 2002) and fuzzy models (Sabater and Sierra, 2001). Most of these algorithms are proposed for decentralized systems. A number of papers have reviewed and compared those academic algorithms (Jøsang et al., 2007; Mui et al., 2002a; Sabater and Sierra, 2005).

This research concentrates on the algorithms that are used by the commercial world. To give a better illustration, this section describes the most common algorithms below. It should be noted, all these algorithms can be used for target rating ($O(r)$), EC rating ($O(c)$) and feedback rating ($O(d)$). For simplicity's sake, the following section uses the target rating to represent all the other ratings; thus r_i is used to represent all the ratings. 'Evaluator' is used to represent evaluators, EC Providers and Feedback Providers.

The nature of r_i

Evaluators are asked to give ratings (r_i) on the target. Assume R is the set of the ratings a target received from all the evaluators:

$$R = \{r_1, r_2, \dots, r_n\}$$

n is the number of reviews (ratings).

1. Most of the time, reputation systems ask evaluators to give an overall rating (r_i) on the target directly. The rating is selected within a range, $r_i \in \{s_1, s_2, \dots, s_n\}$. Different systems choose to use different rating ranges. For example, Amazon uses the Likert scale for evaluators to rate from $\{1, 2, \dots, 5\}$, whereas, eBay asks evaluator to rate the target with binary ratings $\{-1, 0, 1\}$. Thus, the overall rating made by the i th evaluator:

$$r_i \in \{s_1, s_2, \dots, s_n\}$$

2. However some systems require evaluators to rate the target from a number of dimensions ($d_1, d_2, \dots, d_m : d_m \in \{s_1, s_2, \dots, s_n\}$). Then all the dimensions are aggregated with different weights (w_{rj}) to a final rating (r_i). Therefore, the final overall rating made by the i th evaluator can be calculated by:

$$r_i = \frac{1}{m} \sum_{j=1}^m d_j * w_{rj} \quad (4.18)$$

In this case, $s_1 \leq r_i \leq s_n$.

The aggregation algorithms

Most reputation systems choose to use simple central tendency algorithms to compute the overall ratings.

- **SUM:** summation algorithm. This algorithm is widely used to aggregate binary ratings, i.e., $\{-1, 0, 1\}$. The algorithm can be represented by Equation (4.19):

$$S(r) = \sum_{i=1}^n r_i \quad (4.19)$$

Because the algorithm simply adds the ratings, the result may keep growing. In other words, the result can be a very large number. However it is ambiguous to say that the larger the $S(r)$, the better the target. For instance, there are two targets which have the same overall rating, say, 500. The first one (Alice) earns 600 positive points with 100 negative points, whereas, the other (Bob)

gets 500 positive points without any negative one. It is obvious that Bob has a better reputation than Alice, however it cannot be presented by the *SUM* algorithm. Therefore, if a system chooses to use *SUM*, it must employ other algorithms for support.

- *PCT*: percentage model, which calculates the percentage of specific ratings. Assume, a reputation system uses simple $r_i \in \{s_1, s_2, \dots, s_n\}$, R_{s_n} is the set of s_n ratings. Then the set of all ratings (R) consists of the sets of different ratings,

$$R = \{R_{s_1}, R_{s_2}, \dots, R_{s_n}\}$$

Then the proportion of s_k is:

$$P(r) = \frac{|R_{s_k}|}{n} \quad (4.20)$$

Most of the time, *PCT* is used to assist with other algorithms. Back to the example discussed in the previous section, Alice and Bob have the same overall ratings (500), the percentage of positive ratings of Alice is $600/(600 + 100) = 85.71\%$. The percentage of positive ratings of Bob is 100%. $P(r)$ shows that Bob has a better reputation than Alice.

The algorithm can also be adapted to estimate the percentage of a combination of ratings. For example, Amazon's Marketplace allows evaluators to rate the target from $\{1, 2, \dots, 5\}$, within which, 1, 2 are considered as negative ratings, and 4, 5 are positive ones. Then the system calculates the total proportion of positive ratings (4 and 5).

- *AVG*: arithmetic mean algorithm. This algorithm is usually adopted by the systems that use the Likert Scale, e.g., evaluators are allowed to choose ratings from 1 to 5.

$$A(r) = \frac{1}{n} \sum_{i=1}^n r_i \quad (4.21)$$

Unlike the *SUM*, $A(r)$ is a number within the original range, i.e., $\{s_1 \leq A(r) \leq s_n\}$. In other words, for a 5-level rating scale, a target with a 3.5 is considered to have a worse reputation than the one with a 4.5 score.

- *WA*: weighted average algorithm. Some systems give weights to different evaluators (w_{ei}). Then,

$$W(r) = \frac{1}{n} \sum_{i=1}^n r_i * w_{ei} \quad (4.22)$$

The above four algorithms are the most commonly used algorithms. Occasionally, systems may use other algorithms to aggregate ratings, such as, median or mode.

4.3.2 Attributes of Algorithms

4.3.2.1 Robustness

Criterion P 4. *Algorithm Robustness*

This criterion is defined to assess the robustness of the three algorithms. According to basic statistical analysis, one simple way to measure the robustness is to check the *breakdown point* of the algorithm. The breakdown point (ϵ) is the proportion of manipulated ratings required to make the algorithm return an arbitrary value. A higher breakdown point indicates a more robust algorithm (Lewis-beck, 1993). For example, assume R is the set of ratings of a target, $R \in \{r_1, r_2, \dots, r_n\}$. The overall rating of the target is R_t . If someone wants to change the value of R_t , They need to add a number of manipulated ratings (m) to R . Then, $\epsilon = \frac{m}{n}$. Therefore, the larger the ϵ , the more robust the algorithm is. It can be found that *SUM*, *AVG* and *PCT* have very low robustness ($\epsilon = 1/n$), whereas median and mode have relatively high robustness.

If an algorithm is robust, the new ratings cannot easily change the results of the overall rating. In other words, a robust algorithm is not sensitive to the change of new ratings. Reputation systems need to find a balance between the robustness and sensitiveness of the algorithms.

4.3.2.2 Update Frequency

Criterion P 5. *Update Frequency*

Update frequency refers to how often a system updates its ratings and other information. In other words, it assesses how often the algorithms run. Some systems update their ratings as soon as a new evaluator submits the information, while others may update information on a daily or weekly basis.

4.3.2.3 Complexity

There are two kinds of complexities of the reputation systems: algorithm complexity and system complexity.

Criterion P 6. *Algorithm Complexity*

The *algorithm complexity* refers to the complexity of each algorithm, which relates to the analysis of algorithms (Levitin, 2001). As discussed earlier, most centralized reputation systems use very simple algorithms to calculate the ratings, and under the development of current computing technologies, the differences between these algorithms is very little.

Criterion P 7. *System Complexity*

System complexity is the complexity of the whole system, which can be measured as the number of features that a system can provide. For instance, if a system only provides one overall rating without any feedback loop results nor evaluator credibility calculation, it is less complex than the one provides all the three aggregations.

In addition to rating algorithms, processing can also provide other functions, such as allowing users to filter reviews. The more functions a system provides, the more complex the system is.

4.4 Output

From an information system perspective, *output* is the production of useful information, usually in the form of documents and reports (Stair et al., 2010). In terms

of reputation systems, it refers to how the reputation information is disseminated and presented to the end users.

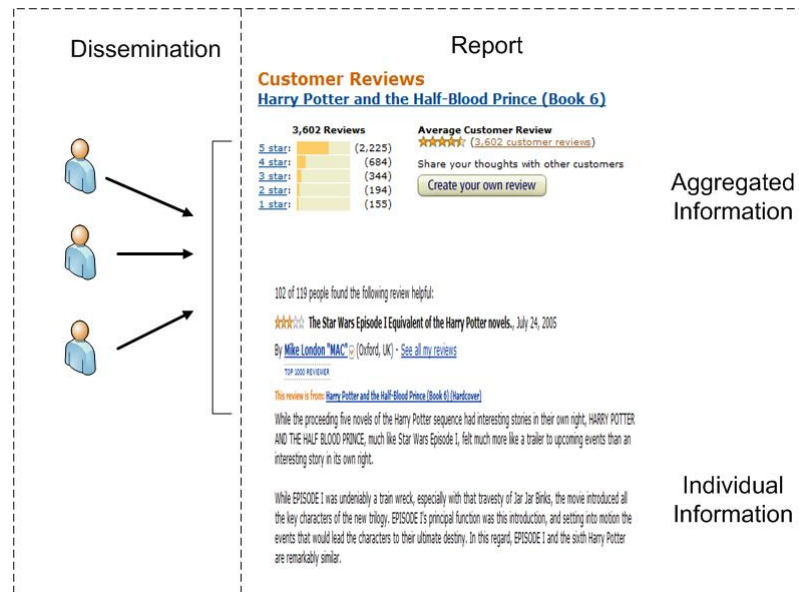


Figure 4.7: Output Criteria

The dissemination indicates how the information is disseminated to the end users (*dissemination* in Figure 4.7). There are two main kinds of information that a reputation system needs to report: *aggregated information* and *individual information*. The former shows the results of the *processing* component, such as the overall rating, and the latter presents the individual ratings and reviews that are collected through the input component (the right side of Figure 4.7 shows the screen shots of the aggregated and individual information of Amazon.com). Reputation systems need to present individual information with details, while aggregated information must provide concise and comparable results to users, so that users can have a clear impression of the target.

4.4.1 Dissemination

Criterion O 1. *The Set of End Users*

The set of end users refers to who are the end users, i.e., who can retrieve the output information. Systems can allow all their site visitors (U_v) to access their published reputation information or require end users to register with them first

(U_r). Reputation systems can also reserve some information for restricted users (U_t). For example, IMDb, the Internet movie database, shows the Top 250 movies to all visitors, but offers the Top 500 to their IMDbPro users, who have paid subscription fees.

Criterion O 2. *Dissemination Methods*

Unlike distributed systems, all centralized reputation systems publish information on their websites. Therefore, this criterion focuses on whether the system supplies alternative ways for their users to get information. Some systems can send emails to users when a new review has been left for targets of interest. Moreover, systems may provide an RSS (Really Simple Syndication) feed for users to track new reviews of a target or the ones submitted by a specific evaluator.

Dissemination methods have two roles: 1) it can help users to get their desired information as soon as possible; 2) with a good use of current technologies, users can help systems to disseminate information to others. For example, in recent years, with the booming of social networking sites (SNS), more and more web sites allow users to share information with their friends. If reputation systems can take advantages of these services, it will not only help users to retrieve information but also bring more new users to the sites.

4.4.2 Aggregated Information

The aim of providing aggregated reputation information is to present the target's reputation in a concise and comparable format.

Criterion O 3. *Timeliness*

Sometimes the target's quality may change over time, for example, a hotel may provide better room services than it used to. Therefore, it is important for reputation systems to be able to present the target's overall ratings in different time periods. For instance, eBay presents the seller's overall ratings during the last month, 6-months and 12-months (Figure 4.8).

Criterion O 4. *Descriptive Dimensions*




| Recent Feedback Ratings (last 12 months) ? | | | |
|--|---------|----------|-----------|
| | 1 month | 6 months | 12 months |
|  Positive | 10734 | 35672 | 67327 |
|  Neutral | 76 | 317 | 536 |
|  Negative | 126 | 365 | 652 |

Figure 4.8: Ratings Presented in Different Time Periods

It is essential that reputation systems can present aggregated information from different aspects. Thus this criterion specifies how many dimensions a system uses to illustrate the target's aggregated reputation information. Usually ratings can be presented by the following three ways: 1) Overall ratings, which are aggregated by the processing algorithms; 2) Rating distributions refer to whether the system can illustrate the distribution of ratings, such as, how many ratings go for 4 in a 1-5 Likert scale. Furthermore, users may want to know the demographic distribution of the ratings, so that they can find the opinions of people who are similar to themselves. 3) Other relevant information, such as the number of total reviews/ratings or other basic statistics data.

Text reviews are not as easy to aggregate as ratings, but systems can use semantic analysis programmes to generate some common factors from reviews and publish the information as review highlights.

It is obvious that the more dimensions the system can provide, the more information the end users can get. However, it also should be noted that more dimensions may increase the system's complexity (Criterion *P6*).

4.4.3 Individual Information

Individual information is the information that is provided by each evaluator. It is necessary for reputation systems to present the raw information as it collected. In addition, reputation systems also need to provide more information on the evaluator and the feedback of the reviews. Moreover, with the growing number of reviews,

reputation systems must consider how to help end users find their desired reviews more efficiently.

Criterion O 5. *Information filtering*

As noted in Section 4.2.2, reputation systems need a sufficient number of reviews to represent the true reputation of the target. However when more and more reviews come out, information overload can occur. Therefore, reputation systems must provide information filtering functions to help users to find desired reviews more efficiently. Information filtering specifies the filtering function reputation systems can provide to the users. It includes filtering and sorting reviews by different dimensions, such as, the given rating and the date the review is left.

Criterion O 6. *Evaluator Information*

In the real world, a person's identity and their personal character can affect trust (Nissenbaum, 2001). Therefore, reputation systems need provide information about the evaluators, for example, their rating/review histories, credibility (if applicable) and even their real names. End users then can use the information to make decisions.

Criterion O 7. *Feedback Loop Information*

This criterion identifies how feedback loop information is published. When presenting the feedback loop results, reputation systems need to provide the full results (e.g., the number of helpful votes and the number of unhelpful votes) or merely the number of helpful votes. Section 4.5 has more discussion on the other features of the feedback loop.

4.4.4 Response Time

When evaluating information systems, response time is always a vital factor. It assesses how quickly a system reacts to users' inquiries. However as discussed earlier, reputation systems do not exist in isolation. They are integrated with web sites. It is difficult to measure their response time without discussing the sites' features. Therefore, this research does not list response time as a criterion.

4.5 Feedback Loop

The quality of the review determines whether a reputation system can work properly. One of the best ways to control the quality is to let users to assess the reviews. In other words, a feedback loop works as a simple version reputation system in which the targets are the reviews. Thus, a feedback loop can also be divided into ‘input, processing and output’ components.

- The input of the feedback loop is the collection of feedbacks. This part can be evaluated in a similar fashion as the input component.
- The processing of the feedback has two meanings: calculation algorithms, which have been defined as Criterion *P3* in Section 4.3.1, and the function (or roles) of the feedback loop.
- The output of feedback is to publish the feedback results, which have been measured in the output component.

Therefore, the criteria of the feedback loop can be grouped into the feedback function and the feedback collection.

4.5.1 Feedback Function

Criterion F 1. *Feedback Loop Function*

The feedback loop function refers to the roles of the feedback loop. The aim of the feedback loop is to assess the quality of the reviews. In other words, the major role of feedback loop is to detect review spams. As discussed in Section 2.4.2, there are two main kinds of review spams: Untruthful reviews and Non-reviews. Thus, the functions of feedback loop are:

Function 1: Reputation systems allow users to rate the reviews as ‘helpful’ or ‘not helpful’ (Figure 4.9a) to identify the untruthful reviews. Systems can also use the results of this kind of feedback to rank the reviews, so that end users will see the most helpful review first.

Function 2: Non-reviews can be deterred by allowing users to ‘report’ or ‘flag’ them (Figure 4.9b). It is understandable that a reputation system may receive a lot of improper information, such as advertisements.

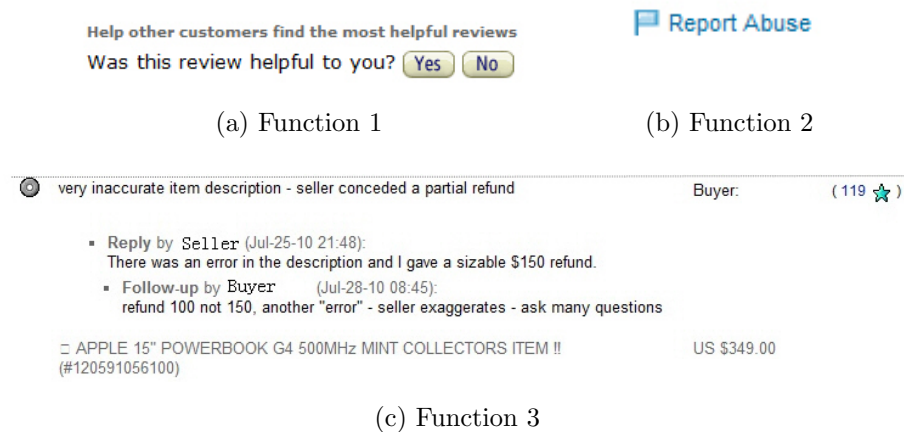


Figure 4.9: Different Functions of Feedback Loop

In addition to identifying the review spams, the feedback loop can also be used to provide more information.

Function 3: Usually, there is no need to worry about whether a single review covers all the aspects of the target, as long as there are sufficient reviews. However in some special cases, such as eBay, if a buyer leaves a negative review of the seller, it would be unfair if the seller does not have the opportunity to provide information from their perspectives. Therefore, the third function of the feedback loop is to provide more information. Figure 4.9c is the snapshot of a eBay’s feedback page. It showed that the seller and buyer have provided more information in the feedback loop.

4.5.2 Feedback Collection

The collection of feedback is much simpler than that of Input. Nearly all feedback is collected through the web page directly (Collection Channel C_{1a} , in Section 4.2.1). Granularity and the providers credibility are also less important because the target

is the review. Thus, the research will not discuss the collection channel, granularity and credibility of the feedback loop.

Criterion F 2. *The set of Feedback Provider*

This criterion defines the set of the feedback providers. Similar to the discussions on Criterion *I2*, in Section 4.2.2, the size of the set of feedback providers (U_{fe}) can be calculated by:

$$|U_{fe}| = |U_{fq}| * p_{fe} \quad (4.23)$$

U_{fq} is the set of eligible feedback providers and p_{fe} is the proportion of providers who actually leave feedback. As feedback are all collected from system's own site, the first three sets of U_q can be applied to U_{fq} : all system visitors can leave feedback (U_v , *Set1* in Equation (4.24)), only registered users can be feedback providers (U_r , *Set2* in Equation (4.24)) and the registered users who are qualified to further restrictions (U_t , *Set3* in Equation (4.24)). Because feedback is usually collected from own site users, thus the *Set4* and *5* in Criteria *I2* are not applicable for the feedback providers.

$$|U_{fe}| = \begin{cases} |U_v| * p_{fe} & \text{Set 1: if all system visitors can be feedback evaluators} \\ |U_r| * p_{fe} & \text{Set 2: if only registered users can leave feedback} \\ |U_t| * p_{fe} & \text{Set 3: if only people have registered with the site} \\ & \text{and are qualified for further restrictions can leave feedback} \\ 1 & \text{Set 4: only one provider is allowed} \end{cases} \quad (4.24)$$

There is a special case in the *Set3*: only one feedback provider is allowed. As discussed in the previous section, when reputation systems use feedback loop for Function 3, they usually only allow one provider. For example, a C2C system usually only allows the seller to leave feedback to the buyer's review. In this case, the size of this set of feedback provider is 1 (*Set4* in Equation (4.24)).

Criterion F 3. *Feedback Format and Breadth*

Similar to input format and breadth, this criterion refers to the format of feedback information and how many properties are collected. As discussed earlier, feedback loop is a simple version of input, usually it only collects numeric ratings and text feedback. The breadth thus refers to how many ratings and text feedback are collected.

Criterion F 4. *Feedback Loop Level*

Sometimes reputation systems allow people to reply to the feedback, in other words, users can leave multiple level feedback. For example, Figure 4.10 shows that Reddit, an online information centre, allows 5 feedback levels. The first feedback was the reply made by ‘kushari’ to the comment of ‘camalex’. Then ‘darth_brooks’ and ‘kushari’ kept replying to each other’s comments for 2 more levels.

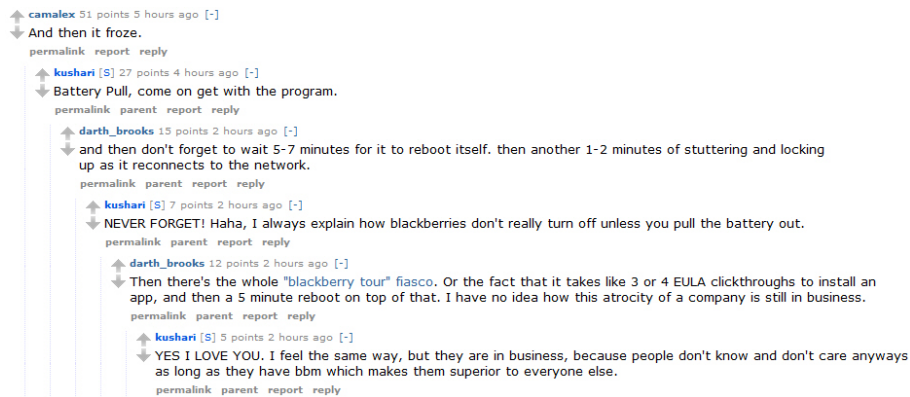


Figure 4.10: Reddit’s Multi-Level Feedback

Multiple feedback levels can provide users more opportunities to discuss the details of the targets.

4.5.3 Feedback Loop Collection Cost

Criterion F 5. *Feedback Loop Collection Costs* (T_{fd})

The collection costs of feedback loop (T_{fd}) is similar to the input collection costs (Criterion I7). It indicates the time it takes a provider to leave a feedback. As indicated earlier, feedback loop only use channel C_{1a} to collect feedback. Thus, similarly to the analysis of T_{ip} which discussed in Section 4.2.4, feedback providers

first need to locate the page they can leave feedback then complete and submit the information. In other words, T_{fd} can be calculated as the same as $T_{ip,1a}$:

$$T_{fd} = T_{br} + T_{cp} \quad (4.25)$$

Therefore, according to the analysis of Equation (4.3) – (4.11):

$$T_{fd} = 67.6 + 1.2 * N_{ff1} + 3.16 * \sum_{i=1}^{N_{ff2}} W_{pf,i} \quad (4.26)$$

N_{ff1} and N_{ff2} denote to the number of ratings and text feedback and the $W_{pf,i}$ is the number of words of the i th text feedback.

4.6 Storage

After the collection and aggregation, information needs to be stored. As discussed in Section 1.1, reputation systems can store the reputation information at either centralized or distributed locations. Because this research focuses on the centralized systems only, there is no need to specify the storage location.

Other possible storage measurements, such as data storage speed and capacity, are usually associated with the hardware and software that have been selected by the web sites. Again, these are out of the research's scope. Therefore, the evaluation of storage concentrates on the storage costs only.

There are three kinds of information that need to be stored: information collected from input, information collected from the feedback loop and the information generated by processing. All three costs can be measured by the data size (the number of bytes). It should be noted that all the storage costs are estimated based on the information of one target, e.g., the input storage cost is the size of the information of one target.

4.6.1 Input Information Storage Cost

Criterion S 1. *Input Storage Cost* (S_{ip})

The size of the input information (S_{ip}) can be estimated by the summation of the size of information provided by each evaluator ($S_{pr,i}$):

$$S_{ip} = \sum_{i=1}^{N_{tr}} S_{pr,i} \quad (4.27)$$

N_{tr} is the number of total reviews, which has been discussed in Section 4.2.2.

The size of the data is related to the format of the reputation information (Criterion $I5$). For example, one rating, which is usually represented by a number, only occupies one byte of storage, whereas a 100-word text review may require more than 600 bytes. A picture or video needs considerably more storage spaces. The breadth of collected information (Criterion $I6$), also has influences on the size of the information. Then,

$$S_{pr} = N_{if1} * S_{if1} + \sum_{j=1}^{N_{if2}} S_{if2,j} + \sum_{j=1}^{N_{if3}} S_{if3,j} \quad (4.28)$$

S_{if1} , S_{if2} and S_{if3} denote the size (bytes) of ratings, text comments and rich media format information an evaluator has submitted. N_{if1} , N_{if2} and N_{if3} refers the number of the corresponding format.

4.6.1.1 The Size of Different Format Information

- S_{if1} : The size of a rating depends on the number of choices (n), $S_{if1} = \log_2(n-1)$. Thus, unless there are more than 256 choices, the size of single rating is less than 1 byte. For practical purposes,

$$S_{if1} = 1 \quad (4.29)$$

- S_{if2} : The size of a text review is decided by the number of characters of each text review (C_{pr}). If the review has 100 characters, then $S_{if2} = 100$. Therefore,

$$S_{if2} = C_{pr} \quad (4.30)$$

According to Shannon (1951) the average word length of English is 4.5 letters, plus the space between two words, then

$$S_{if2} = 4.5 * W_{pr} + (W_{pr} - 1) = 5.5 * W_{pr} - 1 \quad (4.31)$$

- S_{if3} : The size of rich media format information. It is difficult to give a general estimation on the size of a picture or a video. Because there are many factors can influence it, such as different file formats and the quality of the file. For example, for the same image, the JPG format is supposed to have a much smaller size than the PNG format. Thus the research will not discuss on the details on how to generalize the value of S_{if3} .

Therefore, according to above analysis:

$$\begin{aligned} S_{pr} &= N_{if1} * S_{if1} + \sum_{j=1}^{N_{if2}} S_{if2,j} + \sum_{j=1}^{N_{if3}} S_{if3,j} \\ &= N_{if1} + \sum_{j=1}^{N_{if2}} (5.5 * W_{pr,j} - 1) + \sum_{j=1}^{N_{if3}} S_{if3,j} \\ &= N_{if1} + 5.5 * \sum_{j=1}^{N_{if2}} W_{pr,j} - N_{if2} + \sum_{j=1}^{N_{if3}} S_{if3,j} \end{aligned} \quad (4.32)$$

4.6.1.2 Total input storage costs

According to Equation(4.27)-(4.32), the total input storage cost can be calculated as:

$$S_{ip} = \sum_{i=1}^{N_{tr}} (N_{if1} + 5.5 * \sum_{j=1}^{N_{if2}} W_{pr,j} - N_{if2} + \sum_{j=1}^{N_{if3,i}} S_{if3,j}),i \quad (4.33)$$

4.6.2 Feedback Storage Costs

Criterion S 2. Feedback Storage Costs (S_{fd})

The feedback storage is the summation of all feedback collected from all the providers and levels:

$$S_{fd} = \sum_{i=1}^L \sum_{j=1}^{|U_{fe}|,i} S_{pf,ij} \quad (4.34)$$

L is the number of feedback level, $|U_{fe}|$ is the number of feedback providers. S_{pf} is the total size of the information that a feedback provider submits. It is related to the feedback format and breadth (Criterion F3):

$$S_{pf} = N_{ff1} * S_{ff1} + \sum_{k=1}^{N_{ff2}} S_{ff2,k} \quad (4.35)$$

S_{ff1} and S_{ff2} denote the size (bytes) of ratings and text comments that a provider submits as feedback information. N_{ff1} and N_{ff2} refer the number of the corresponding format.

Therefore, according to the analysis of Equation (4.29)–(4.32):

$$S_{pf} = N_{ff1} + 5.5 * \sum_{k=1}^{N_{ff2}} W_{pf,k} - N_{ff2} \quad (4.36)$$

W_{pf} is the words for each text feedback. Equation (4.34) then becomes:

$$\begin{aligned} S_{fd} &= \sum_{i=1}^L \sum_{j=1}^{|U_{fe}|,i} S_{pf,ij} \\ &= \sum_{i=1}^L \sum_{j=1}^{|U_{fe}|,i} (N_{ff1,ij} + 5.5 * (\sum_{k=1}^{N_{ff2}} W_{pf,k})_{,ij} - N_{ff2,ij}) \end{aligned} \quad (4.37)$$

4.6.3 Processing Information Storage Costs

Criterion S 3. *Processing Information Storage Costs* (S_p)

The *processing information* refers to the information that is produced during the processing. It includes any information that is generated by the processing. The storage cost of processing information can be estimated as the summation of all kinds of information:

$$S_p = N_{pp1} * S_{pp1} + \sum_{j=1}^{N_{pp2}} S_{pp2,i} + \sum_{i=1}^{N_{pp3}} S_{pp3,i} \quad (4.38)$$

S_{pp1} , S_{pp2} , S_{pp3} denote to the size of ratings, text information and rich media format information that have been produced. N_{pp1} , N_{pp2} , N_{pp3} are the number of corresponding information formats.

Therefore, according to the analysis of Equation (4.29)–(4.32):

$$S_p = N_{pp1} + 5.5 * \sum_{i=1}^{N_{pp2}} W_{pg,i} - N_{pp2} + \sum_{i=1}^{N_{pp3,i}} S_{pp3,i} \quad (4.39)$$

$W_{pg,i}$ is the words for the i th text information.

4.7 Conclusion

The SERS model systematically analyzing reputation systems from the underlying structure perspective, from which all systems can be divided into five components: input, processing, output, feedback loop and storage. In total, 29 criteria have been defined based on the five components. There are 7 criteria for the Input, 7 for the Processing, 7 for the Output, 5 for the Feedback Loop and 3 for the Storage.

These criteria can be classified into three groups (Table 4.1):

1. **Classification criteria** can show the differences between reputation systems, i.e., these criteria can distinguish reputation systems from one to another. However the measurement of this kind of criteria are not comparable. Take collection channel (Criterion I1) as an example. It cannot simply say that a system uses C_2 is better than the one uses C_{1a} .
2. **Measurement criteria** assess the performance of reputation systems from different aspects. For example, if a system provides more descriptive dimensions (O4), it is better than the one provides less dimensions.
3. **Cost criteria**. As discussed in Section 2.6, the cost of reputation systems have been long ignored, thus this research distinguish the criteria, which focuses on the costs specifically, from the measurement criteria. For example, input and feedback collection costs concentrate on the time costs for collecting information.

| | Classification Criteria | Measurement Criteria | Cost Criteria |
|---|---|--|--|
| I | I1. Collection Channel I5. Information Format | I2. The Set of Evaluators I3. Granularity I4. Evaluator Credibility I6. Breadth | I7. Input Collection Cost |
| P | P1. Rating Algorithm P2. Evaluator Credibility Algorithm P3. Feedback Aggregation Algorithm | P4. Algorithm Robustness P5. Update Frequency | P6. Algorithm Complexity P7. System Complexity |
| O | O2. Dissemination Method | O1. The Set of End Users O3. Timeliness O4. Descriptive Dimensions O5. Information Filtering O6. Evaluator Information O7. Feedback Information | |
| F | F1. Feedback Loop Function | F2. The Set of Feedback Providers F3. Format and Breadth F4. Feedback Loop Level | F5. Feedback Collection Cost |
| S | | | S1. Input Storage Cost S2. Feedback Storage Cost S3. Processing Information Storage Cost |

Table 4.1: Criteria Summary

Chapter 5

Theoretical Evaluation of the SERS

5.1 Introduction

This research proposes the SERS (Systematic Evaluation of Reputation System) model for measuring the intrinsic characteristics of centralized reputation systems. This chapter evaluates the SERS from the theoretical perspective.

By far the SERS is the first model that aims at systematically measuring different types of reputation systems; thus there is no similar model that it can be compared against. As an alternative, Section 5.2 discusses whether the SERS has identified the major successful factors of reputation systems, which are proposed by relevant literatures.

Moreover, as reputation systems are essentially information systems (IS), Section 5.3 then compares the SERS with the IS evaluation dimensions.

5.2 Influential Factors

Section 2.4 has surveyed the literature surrounding the evaluation of reputation systems. Based on this discussion, Table 2.1 listed a number of characteristics that influence the success of reputation systems. Table 5.1 re-lists the characteristics and compares them with the criteria defined in the SERS.

| Characteristics | | Corresponding SERS Criteria |
|---------------------------------|--------------------|-----------------------------|
| Source | Sufficient Sources | <i>I2</i> |
| | Credibility | <i>I4</i> |
| | Granularity | <i>I3</i> |
| | Type | <i>O6</i> |
| Information Quality | | see Section 5.3.1 |
| Information Format | | <i>I5, F3</i> |
| Information Aggregation | | <i>P1, P2, P3, O4</i> |
| Time-related | Timeliness | <i>O3</i> |
| | Update Frequency | <i>P5</i> |
| Information Filtering Mechanism | | <i>O5</i> |

Table 5.1: The SERS vs the Successful Factors of Reputation Systems

In general, most characteristics proposed by the literatures have been directly defined in the SERS model, such as *sufficient sources*, *credibility*, *granularity*, *information format*, *information aggregation*, *timeliness*, *update frequency* and *information filtering mechanism*.

Although the SERS model did not identify the ‘type of sources’, one of the *Output* criteria, *O6* (Evaluator Information) assesses whether systems can provide information of evaluators types to the end users.

There is no criterion has been assigned to directly measure the quality of the reviews; however, the quality of reviews can be identified by the value of a number of relevant criteria. More discussion can be found in Section 5.3.1, which discusses the accuracy of reputation information.

5.3 IS Dimensions VS the SERS Criteria

As indicated in Section 2.5, reputation systems are information systems that use the Internet as an intermediary. This section compares the SERS model with IS evaluation models.

Section 2.5 had a full review of IS literature and discussed a number of major

measurement dimensions. Table 5.2, which reproduced from Table 2.2, is the list of relevant dimensions of Information Quality (IQ) and System Quality (SQ). IQ and SQ are the basic technical quality measures of IS.

| IQ Dimensions | SQ Dimensions |
|----------------------|----------------------|
| Accuracy | Usability |
| Completeness | Reliability |
| Timeliness | Response Time |
| Accessibility | Usefulness |
| Interpretability | |

Table 5.2: IQ and SQ Dimensions

The following sections discuss the correlation between the SERS criteria and the IS dimensions. There are three correlations between the criteria and dimensions: 1) criteria conform to the dimensions completely, if they have the same meaning or measuring the same attributes; 2) criteria can reflect the value of the dimensions, or can be used as the measurement of the dimensions; 3) a dimension may have no correlation with any criterion.

5.3.1 IQ Dimensions

The information quality of reputation systems refers to the quality of the reputation information.

5.3.1.1 Accuracy

No criterion conforms to this dimension. However, a number of criteria can be used to reflect the accuracy of reputation information. The accuracy of the reputation information can be specified in two aspects:

1. The accuracy of individual reputation information.

As discussed in Section 2.4.5, as long as evaluators tell their true opinions about targets, their ratings and reviews are accurate. In other words, review spams

are considered as the inaccurate reviews. Thus, the accuracy of individual information is related to the factors that can reduce or filter review spams.

The reliability and credibility of evaluators can be used to assess the quality of the information (Smith, 1997; Huang et al., 2010; Pernici and Scannapieco, 2002). In the SERS model, *I4* (Evaluator Credibility) and *P2* (Evaluator Credibility Algorithm) are the two criteria that represent the property of evaluator credibility. Moreover, Criterion *I3* (Evaluator Granularity), which identifies how an evaluator is related to the target, can partially reflect the reliability of the evaluators.

In addition, with the help of the feedback loop, reputation systems can identify the helpfulness of the reviews or can identify spam information. Therefore, *Feedback Loop* Criteria *F1* – *F4* can be seen as measurement of accuracy.

2. The accuracy of aggregated reputation information refers to whether the overall published information can reflect the true quality of the target.

The accuracy of aggregated reputation information depends on two main factors: the aggregation algorithms and the number of evaluators. Processing algorithms Criteria *P1*, *P2* and *P3* specify how the information is aggregated, which are the measurements of the accuracy. As indicated in Section 4.2.2, when there are sufficient number of reviews, the reputation information can reflect the quality of the target. Thus criterion *I2* is also a measurement of accuracy.

In general, due to the nature of reputation systems, accuracy can be reflected by a number of criteria listed in the SERS model.

5.3.1.2 Completeness

Criteria *I6* (information breadth), *F3* (feedback format and breadth) and *F4* (level of feedback) indicate how much information has been collected. These criteria measure the completeness of collected information. *O4* (descriptive dimensions), *O6* (evaluator information) and *O7* (feedback information) measure the completeness of published information.

In addition, *O3* (timeliness) specifies whether systems can show the aggregated information during different time periods, which is partially related to the completeness as well

5.3.1.3 Timeliness

Criterion *O3* conforms to the timeliness completely and the *P5* (update frequency), which assesses how often the information is updated, also conforms to this dimension.

5.3.1.4 Accessibility

Accessibility assesses how users can obtain the information. Two criteria in the SERS model conform to it: *O1* (dissemination method) and *O2* (the set of end users) assess the access methods and the people who can retrieve the information.

5.3.1.5 Interpretability

The interpretability describes how the information is presented to the end users. Output criteria *O4* (descriptive dimensions) and *O5* (information filtering) directly measure the value of the dimension. Furthermore, how the information is collected and aggregated also have great influence on the interpretability. Thus *I5* (information format), *F3* (feedback format) and *P1*, *P2*, *P3*, the aggregation algorithms reflect the value of interpretability as well.

5.3.2 SQ Dimensions

For reputation systems, SQ refers to how good does a system provide their services to the end users.

5.3.2.1 Usability

Usability indicates the ease-of-use of the system. *O5* (information filtering criterion) refers to whether end users can filter reviews as the way they prefer. It can be seen as a measurement of usability.

5.3.2.2 Reliability

Reliability specifies the level of stability of a system and the rate of system failure and error. From the technical perspective, the SERS does not have any criterion relates to reliability. Because most of those features are controlled by the web sites rather than by the reputation systems. However, for reputation systems, the rate of system error also can be seen as the rate of review spams, in particular, non-review spams. End users will lose faith on the system if it collects too many non-review spams, such as advertisement and promotion of other sites.

Criterion $F1$, which indicates the function of feedback loop can be used to assess whether a system has a good spam filtering mechanism. This criterion can be seen as a measurement of reliability.

5.3.2.3 Response Time

As discussed in Section 4.4.4, the response time of reputation systems are decided by the web sites. To concentrate on the intrinsic nature of reputation systems, the SERS does not include response time as a criterion.

5.3.2.4 Usefulness

According to Section 2.5.2.4, the usefulness of reputation systems can be seen as whether the systems can function well and whether the system has the ability to accept customer feedback. Whether a reputation system can function well depends on a number of successful factors, which have been discussed in Section 2.4.5. Section 5.3 has analyzed how the SERS match with these factors, thus there is no need to repeat it.

The customer feedback capability of reputation systems can be measured by $F1$ (feedback function), $F2$ (the set of feedback providers), $F3$ (feedback format and breadth) and $F4$ (feedback level).

5.3.3 Summary

Table 5.3 summaries the correlation between IS dimensions and SERS criteria. The ‘C’ indicates that the criterion conforms to the IS dimension completely, while an ‘M’ means the criterion can be used as a measurement of the dimension.

It can be seen that that the SERS has defined several criteria that conform to the timeliness and accessibility completely. In addition, all the other IQ dimensions and three of SQ dimensions can be measured by some of the criteria. The only criterion that does not reflected by the SERS is response time. As discussed earlier, that is due to the response time is in control of the web sites rather than the reputation systems.

The table also shows that the SERS has defined a number of criteria that are not covered by other literature. For example, *I1* (input collection channel), *P4* (algorithm robustness) and most of the Cost Criteria.

5.4 Conclusion

This chapter assesses the SERS model from the theoretical perspective. It demonstrated that the SERS has identified major influential characteristics of reputation systems, including factors related to the ratings and reviews, information sources and other relevant ones.

By comparing the SERS with IS evaluation models, it showed the SERS has covered most IQ and SQ dimensions. The response time is not reflected because the SERS model only focuses on the intrinsic nature of reputation systems rather than the business operation of the whole site.

Furthermore, the SERS has proposed more characteristics, which have not been identified by other literatures. For example, aggregation algorithms for evaluators credibility and feedback, feedback functions and a number of costs criteria.

| | Input | | | | | | | Processing | | | | | | | Output | | | | | | | Feedback Loop | | | | | Storage | | | |
|------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|--|
| | I ₁ | I ₂ | I ₃ | I ₄ | I ₅ | I ₆ | I ₇ | P ₁ | P ₂ | P ₃ | P ₄ | P ₅ | P ₆ | P ₇ | O ₁ | O ₂ | O ₃ | O ₄ | O ₅ | O ₆ | O ₇ | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | S ₁ | S ₂ | S ₃ | |
| Accuracy | M | M | M | M | | | | M | M | M | | | | | | | | | | | | | M | M | M | M | | | | |
| Completeness | | | | | | M | | | | | | | | | | | M | M | | M | | | | | M | | | | | |
| Timeliness | | | | | | | | | | | C | | | | | | C | | | | | | | | | | | | | |
| Accessibility | | | | | | | | | | | | | | | C | C | | | | | | | | | | | | | | |
| Interpretability | | | | | M | | | M | M | M | | | | | | | | M | M | | | | | | M | | | | | |
| Usability | | | | | | | | | | | | | | | | | | | M | | | | | | | | | | | |
| Reliability | | | | | | | | | | | | | | | | | | | | | | M | | | | | | | | |
| Response Time | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Usefulness | | | | | | | | | | | | | | | | | | | | | | M | M | M | M | | | | | |

Table 5.3: IS Dimensions and the SERS

Chapter 6

Empirical Evaluation of the SERS

6.1 Introduction

One of the best ways to assess a theoretical model is to apply it to commercial applications. This chapter uses the SERS model to assess 15 commercial reputation systems. The results show that the SERS model can distinguish different systems and compare them.

The assessment was carried out using case study method (Yin, 2003; Simons, 2009). According to Yin (2009), the design of case studies should follow four main procedures: *developing the theory*, *selecting cases*, *collecting data* and *analyzing the data*. For this research, the theory is the SERS model, and the cases are the 15 sites. Section 6.2 introduces how the assessment was conducted, including the selection of cases (sites) and the collection of the data. Section 6.3 analyzes the results derived from the data.

6.2 Research Design

6.2.1 Commercial Sites

The selection of the sites are mainly based on the following two principles:

1. The variety of the types and the nature of the sites. One of the aims of the SERS is to evaluate different types of reputation system in the same context.

Thus the selected sites should cover the main system types.

2. The site's popularity in commercial world. The selected sites should be well-known in the real world, so that they can represent other similar sites.

Based on the principles, 15 sites are selected, which cover the different types of reputation systems. As discussed in Section 2.3.4, based on the functions, reputation systems can be classified into three types: 1) *C2C Systems* adopt reputation systems to build trust among strangers; 2) *Review Centres* use reputation systems to reduce information asymmetry and 3) *Online Communities* filter information by reputation systems.

Table 6.1 lists the sites with their targets. The research selects three C2C systems:

- Amazon (<http://www.amazon.com>) has two different markets. **Amazon Marketplace (Amazon M)** is an eBay-like C2C marketplace but with fixed prices. Buyers leave ratings about the sellers after each transaction. Though the site also allows sellers to rate the buyers, it does not use the buyer ratings to evaluate them¹. The other marketplace is their retailer markets (Amazon R), which will be discussed later.
- **eBay** (<http://www.ebay.com>) is one of the largest online auction marketplaces. It provides a platform for consumers to buy and sell items with each other.
- **Elance** (<http://www.elance.com>), is an internet 'job centre' for freelancers. It builds a platform for employers to find freelancers and vice versa. There are two main kinds of users of the site, the employers, who are looking for people to work for them and the freelancers, who are looking for jobs. Like eBay, after the freelancer has finished the job, the employer rate them.

¹Amazon Marketplace Help Page (2011); 'Seller feedback for buyers.' http://www.amazon.com/gp/help/customer/display.html/ref=hp_lnav_dyn?ie=UTF8&nodeId=200278900; Last Accessed 15 January 2011.

| | Sites | Targets |
|--------------------|----------------------------------|---|
| C2C Systems | Amazon Marketplace (Amazon M) | Sellers, buyers |
| | eBay | Sellers, buyers |
| | Elance | Freelancers |
| Review Centers | Amazon Retailer (Amazon R) | General products |
| | Bizrate | Online shops |
| | Epinions | General Products |
| | Google Shopping | Shops, products |
| | IMDb | Movies |
| | Reevo | General Products |
| | Tripadvisor | Travel related items, such as hotels and restaurants |
| Online Communities | Yelp | Restaurants, clubs, shops, etc |
| | Digg | Stories and comments |
| | Reddit | Stories and comments |
| | Slashdot | User Comments |
| | Yahoo! Answers | Answers |

Table 6.1: Summary of Selected Sites

Nine Review Centres are selected:

- Amazon's Retailer platform (**Amazon R**), is one of the largest online retailers. It sells all kinds of goods from books and clothes to groceries.
- **Bizrate** (<http://www.bizrate.com>) is an online shop review centre. It cooperates with a number of online stores. After a customer makes a transaction through a partner shop, Bizrate contacts them and asks for reviews.
- **Epinions** (<http://www.epinions.com>) is an online product review centre. It

collects reviews from general Internet users. One special feature of Epinions is that the evaluators can make money by writing reviews. The site has two earning systems: Eroyalties Credits and Income Share. The former rewards evaluators on how many times their reviews have been visited and the latter rewards evaluators whose reviews can help other users make decisions (make a buying decisions or avoid a purchase)². However, the formula of how these earnings are calculated is not disclosed.

- **Google Shopping** (<http://www.google.co.uk/products>), also known as Google Product Search, is a price comparison service. It collects ratings and reviews from other reputation systems and publishes full or partially information.
- The Internet Movie Database (**IMDb**) (<http://www.imdb.com>) is a huge collection of movie information. Users are allowed to leave ratings and comments on the movies.
- **Reevo** (<http://www.reevo.com>) is an online product review centre, similar to Bizrate, it collects reviews through their online shop partners. It also allows users to submit reviews via their web site as long as the evaluators can provide proof of purchase.
- **Tripadvisor** (<http://www.tripadvisor.com>) is one of the world's largest travel communities where people can write and read reviews on hotels, restaurants and other related information.
- **Yelp** (<http://www.yelp.com>) is a review centre focusing on local businesses. It allows users to rate local amusements, such as restaurants, clubs and shops.

Online Communities usually utilize reputation systems as their core mechanism to filter information. Four sites are selected for this category.

²Epinions (2011). 'Epinions help page'; <http://www0.epinions.com/help/faq>; Last Accessed 15 January 2011.

- **Digg** (<http://www.digg.com>) is an information centre, as discussed before, it allows users to rate and comment on stories and news.
- **Reddit** (<http://www.reddit.com>) is similar to Digg, which also encourages users to vote on the stories or other information shared on their site.
- **Slashdot** (<http://slashdot.org/>) is a discussion board which focuses on technology news. It has a complicated karma system for calculating the member's score. Unlike Digg and Reddit, the story submitted at Slashdot is selected by the editors, but the comments can be rated.
- **Yahoo! Answers** (<http://answers.yahoo.com/>) is an online Q&A community, where users can ask and answer questions. It uses a reputation system to choose the best answers and filter the spam information.

6.2.2 Data Collection

According to Yin (2009), there are six sources of evidence that are most commonly used in doing case studies: documentation, archival records, interviews, direct observations, participant-observation, and physical artifacts. This research mainly collects data from *documentation*, *direct observations* and *participant-observations*.

Data collected by direct observations. Direct observations refer to collect data from the web pages directly. For example, most of Output component data are published on the web pages; thus they can be collected by direct observations.

Data collected by participant-observations. Participant observations are used to collect Input and Feedback Loop data. Usually after registering with a site, it allows to write reviews or leave feedback. Thus, the Input and Feedback Loop data can be collected.

Data collected by documentation. Sometimes, sites may have strict rules on who can write reviews, which means, it is not able to collect Input or Feedback Loop data by participant-observations method. However, most of the time, reputation systems have detailed explanations of how their reviews are

collected on their help pages. Some even have screen shots as examples. Furthermore, part of Processing data can also be revealed from their documentations.

These three sources provide most data for Input, Output and Feedback Loop component. According to the criteria, the data of Storage can be estimated by the data of other components.

The collection of the Processing component data is the most difficult one. Most sites do not disclose their processing information to the public, such as rating algorithms and update frequencies. Therefore, the research treat the Processing component as a black box, which means the value of Processing criteria are estimated by comparing the Input and Output.

However the value of update frequency ($P5$) and algorithm complexity ($P6$) cannot be achieved by this method. Therefore the results of these two criteria will not be discussed. Furthermore, according to Section 4.6.3, the storage cost of processing information ($S3$) is estimated by the summation of the size of the information generated in the Processing. It is not possible to find out how much information is produced internally by the system. Therefore, there is no discussion for this criterion either.

In summary, of all 29 criteria, only 26 criteria will be used to measure the sites.

6.3 Results

This section discusses the results of assessing the 15 sites by the SERS model.

6.3.1 Input

6.3.1.1 I1 Collection Channel

Section 4.2.1 differentiates the C_{1a} and C_{1b} channels by whether the evaluators actively leaving reviews or the sites send reminders to them. Theoretically if a site, can trace users' transactions, such as a retailer store or C2C marketplace, it then can send emails to the evaluators to remind them to provide reputation information.

In this manner, Amazon M, eBay, Elance and Amazon R are all capable to use both C_{1a} and C_{1b} channel. However, practically, these sites may not send the reminders. For example, although eBay can send emails to remind the buyers to write reviews after the transaction is finished, the site does not supply this service³. Rather, it provides a paid service called ‘Selling Manager Pro’, which allows the sellers to send automatic email reminders to the buyers. These sites do not directly disclose publicly on whether they send the reminders or not; therefore, for fairness sake, the research assumes that all the four sites only use C_{1a} channel (see Figure 6.1).

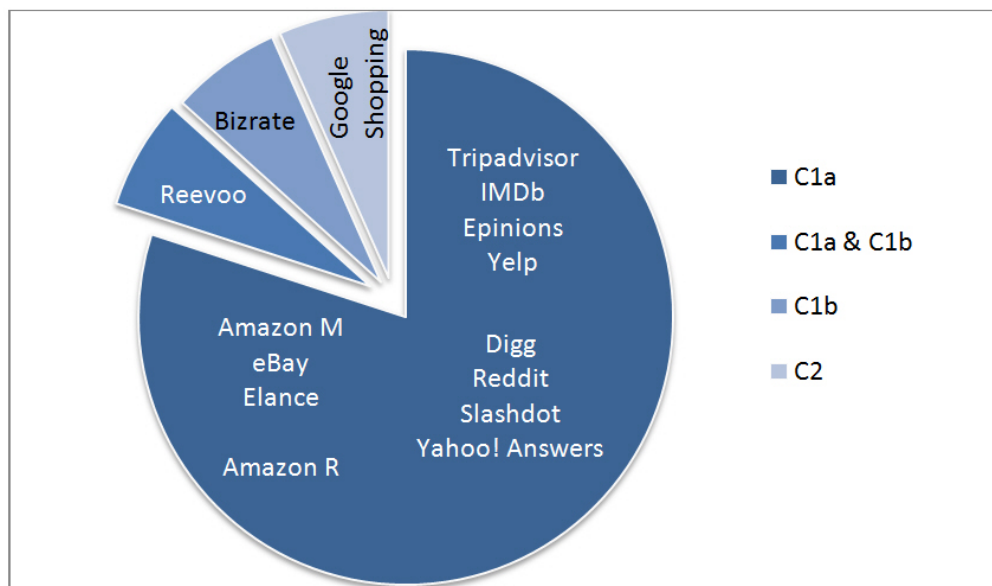


Figure 6.1: Distribution of Collection Channels

Reevo is the only site that can use both C_{1a} and C_{1b} channels. It collects reputation information from their partner shops as well as allows evaluators to leave reviews directly on its website. However due to the site requirements, which ask evaluators to provide proof of purchase, their C_{1a} evaluators are considered to be smaller number than the other sites.

As shown in Figure 6.1, most systems use C_{1a} to collect information, while Google is the only one that utilizes the C_2 channel and Bizrate uses C_{1b} .

³eBay (2011). ‘ebay help page.’ <http://pages.ebay.com/help>; Last Accessed 15 January 2011.

6.3.1.2 I2 Set of Evaluators

The set of evaluators (U_e) measures how many evaluators leave reputation information about the same target. The size of U_e is calculated by $U_q * p_e$ (Section 4.2.2). Practically, it is not possible to assess the p_e of each site, because it is influenced by many site-related factors. Thus, this section assumes all the sites have the same p_e .

Figure 6.2 compares the sets U_e based on the calculation of Equation (4.2). From left to right, the sizes of the sets are becoming larger. That is because, excluding irrelevant influential factors (such as website design), all the sites can be assumed to have the same U_v , then $U_v > U_r > U_t$. Thus, Set 1 > Set 2 > Set 3. Furthermore, the sizes of Set 4 and 5 are larger than Sets 1, 2, 3 because these two kinds of system collect information from more than one site. In addition, Set 5 systems are able to collect information from other reputation systems, including Set 4 systems. Then, Set 5 has the largest number of U_e . In other words, Google shopping has the largest number of U_e among all the systems.

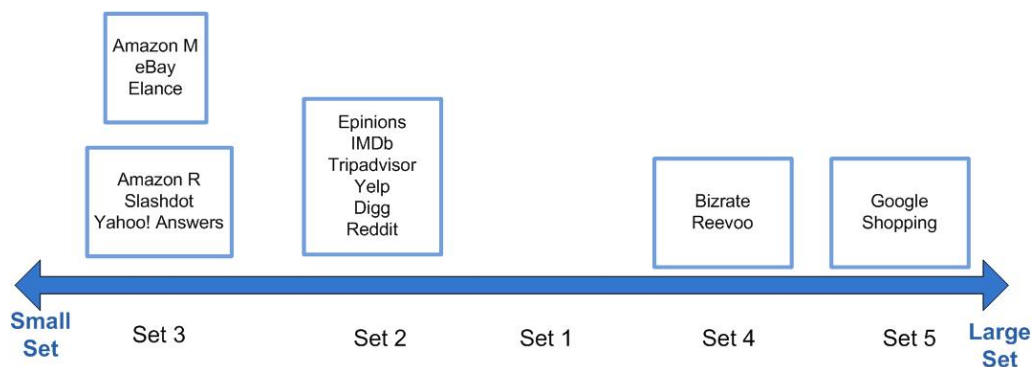


Figure 6.2: Comparison of the Set of Evaluators

It also can be seen from the figure that no site allows people to write reviews without registering with it. 6 sites, including 4 review centres and 2 online communities, accept all registered users to be evaluators. Within the six sites that are in the Set 3, the three C2C systems require their evaluators to have direct transactions with the targets.

Slashdot and Yahoo! Answers require the evaluators to become active users before they can leave reputation information. Both sites adopt the Karma mechanism, which records all the user activities and gives scores to them; thus each user's overall

score (or called Karma) can indicate the level of their activity. To be an evaluator on Amazon R, registered users need to buy at least one item from the site; and after that, evaluators can leave reviews to any products, even the ones they did not buy from Amazon R.

6.3.1.3 I3 Granularity

Most systems do not have good performances on identifying the granularity between the evaluators and targets. 9 out of 15 sites cannot identify neither the interaction nor the expertise granularity (Figure 6.3).

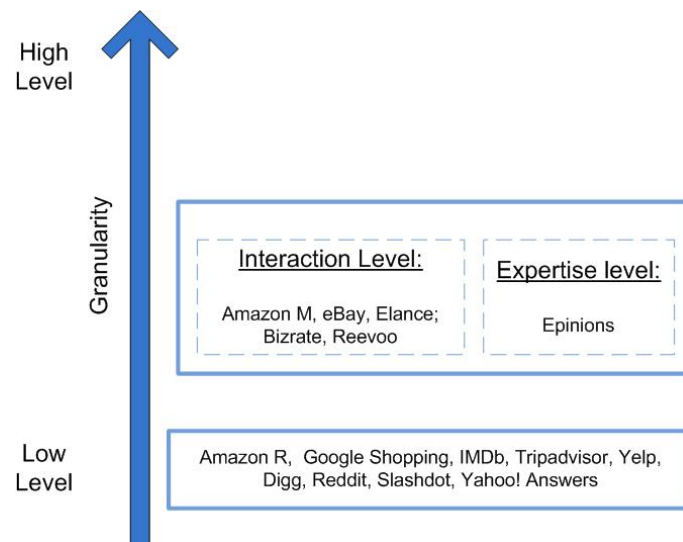


Figure 6.3: Comparison of the Granularity of Evaluators

Reevo, Bizrate and all the C2C systems can identify the interaction between the evaluators and the targets, thus they have interaction granularity. Epinions is the only site that is able to show the expertise granularity of the evaluators. The site selects a number of evaluators according to quantity and quality of the review they have written as ‘Category Leads’ and ‘Top Reviewers’ in each product category⁴. No site has a high granularity that can identify both expertise and interaction granularity.

⁴Epinions (2011). ‘Epinions help page.’ <http://www.epinions.com/help/faq>; Last Accessed 15 January 2011.

6.3.1.4 *I4* Evaluator Credibility

Among the sites that can identify the evaluator credibility:

- Amazon R, Yelp and Reddit take the feedback providers' opinions (results of the feedback loop) to calculate evaluators credibility.
- Epinions is the only site which has two credibility mechanisms, one is controlled by the site (Top Reviewer and Category Lead), the other allows the end users to choose the evaluators they can trust.
- eBay's system allows buyers and sellers to rate each other after the transaction, which means the targets and the evaluators are exchangeable. Thus the evaluators' credibility is decided by the targets.

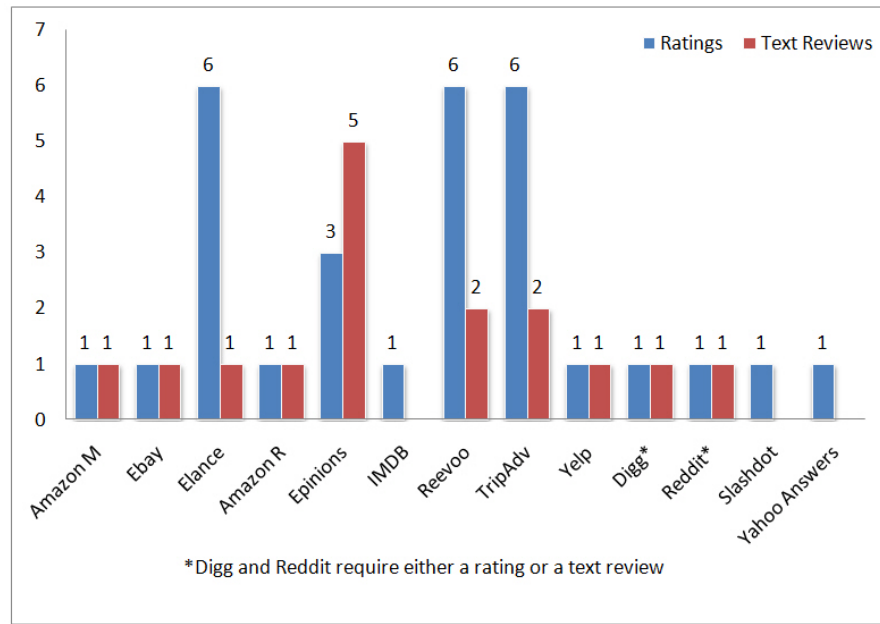
All the other 10 systems do not have any evaluator credibility mechanism at all.

6.3.1.5 *I5* Information Format and *I6* Information Breadth

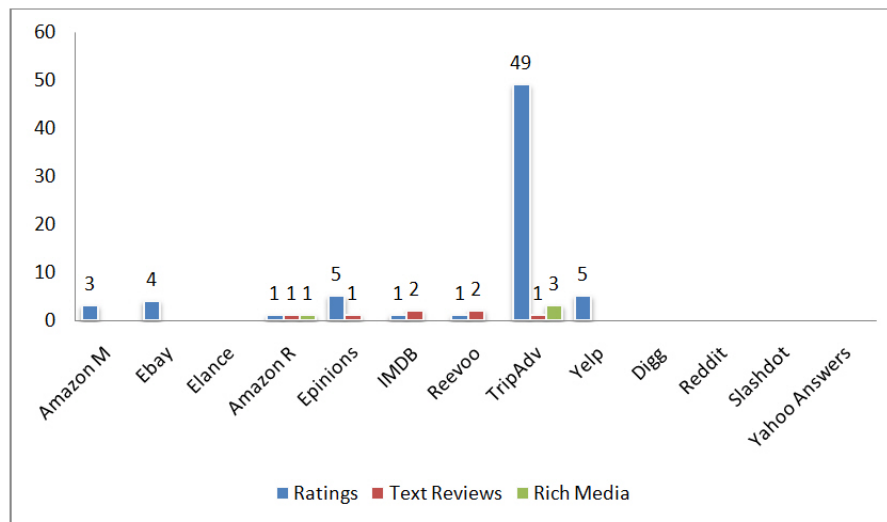
For these two criteria, the data of Bizrate and Google Shopping cannot be revealed, because they collect reputation information from their partner sites. Furthermore in their help pages, they do not make clear on the amount of information they collect. Therefore the following discussion excludes these two sites. It also should be noted that Epinions and Tripadvisor collect different amount of information for different targets. For example, Tripadvisor collect 6 required ratings for hotels, while it only requires 3 ratings for restaurants. The data collected from Tripadvisor are based on their 'hotel' targets and the data collected from Epinions are based on the 'digital cameras'.

As discussed in Section 4.2.3, reputation systems usually mark the information on the collection form as required or optional, so that evaluators can decide how much information they want to provide. Figure 6.4a compares the amount of required information each site collects. It shows most sites require one rating and/or one text reviews.

IMDb, Slashdot and Yahoo! Answers accept as little as one rating, whereas Digg and Reddit require either a rating or a text review. Elance requires 6 ratings



(a) Required Information



(b) Optional Information

Figure 6.4: Information Breadth and Format

because the site uses weighted average algorithm (Equation (4.18) in Section 4.3.1) to calculate the evaluator's overall rating rather than asking for one single rating directly. Reevo and Tripadvisor also ask evaluators to rate the target from multiple quality factors though they also require the evaluator to provide an overall rating. Epinions is the only site that requires more text reviews than ratings. The site aims to provide comprehensive and quality reviews to the end users, thus it requires the

pros and cons and other relevant information from evaluators.

As shown in Figure 6.4b, which compares the optional information, Tripadvisor collects much more optional ratings than the other sites. The site uses check boxes to collect information and asks evaluators to ‘Select All that Apply’. For example, when asking the evaluator to describe their trip, the system provides 18 choices including nightlife, romance and business meetings.

5 sites do not require any optional information at all. The other sites usually ask for more ratings for other quality factors about the target. For example, eBay asks for detailed ratings on ‘Item as described’, ‘Communication’, ‘Dispatch time’ and ‘Postage and packaging charges’.

Rich media format information is not very popular in the systems, Amazon R and Tripadvisor are the only sites that accept it, as optional information only. Amazon R accepts video reviews and Tripadvisor allows pictures.

In general, online communities collect less information than others. One possible reason is that the target of online communities are information, which does not have many quality-factors, while C2C systems and review centres’ targets are services or products, which have multiple dimensions on their quality. For example, service quality can be described as the delivery speed, customer services and quality of the goods. Therefore, the nature of the targets decides the amount of information needs to be collected.

6.3.1.6 *I7* Input Collection Cost

Based on the discussion in Section 4.2.4, the input collection costs are calculated by different collection channels. The collection time for C_2 sites depends on their indexing speed. Google shopping is the only site that uses the C_2 channel. Although it is not possible to have an exact data of Google shopping’s index speed, the number is supposed to be much smaller than the costs of other sites. Because crawling information on the Internet is much faster than filling collection forms by evaluators. In other words, Google Shopping, the only C_2 site, has the lowest input collection cost.

For C_{1a} and C_{1b} systems, the collection costs are essentially decided by the infor-

mation formats and the number of each format. Rich media information requires the longest time to produce and upload, which may cost tens of minutes or even hours to submit. As discussed in the previous section, only two sites accept rich media information (Amazon R and Tripadvisor) and very few evaluators choose to provide rich media information due to the much higher costs. This section concentrates on the collection costs of ratings and text reviews only.

Thus, according to the calculations in Section 4.2.4, the collection costs of C_{1a} and C_{1b} channels can be seen as linear functions of W_{pr} (the words count of the review). Each evaluator may provide different amounts of information when writing reviews. In other words, even in the same system, it takes different amounts of time for evaluators to provide information. This section chooses to compare the sites by three settings: minimum, maximum and 150-word settings.

- **Minimum setting** assumes evaluators provide as least information as they can. For example, they will only make the required ratings and write the required text reviews with 1 word. The aim of this setting is to see how much time it costs an evaluator to provide the least information.
- The intention of **maximum setting** is to estimate how much time it takes an evaluator to provide as much information as possible. In other words, it means an evaluator will give ratings to all the required and optional ones and write all the text reviews with as much words as possible. Some systems have a maximum words limit on their reviews, say, 1000 words. However other systems do not have any words limit, in this case, for convenience sake, this setting uses 2000 words to compare the results.
- Understandably most evaluators are not likely to write a one-word-review nor a maximum-word-review. Therefore the **150-word setting** aims to estimate in a normal circumstance how much time it takes an evaluator to provide information. Research has found that on average there are 150 words per review (Chevalier and Mayzlin, 2006). Thus, this setting assumes evaluators will give ratings to all the required ones and write 150-word reviews.

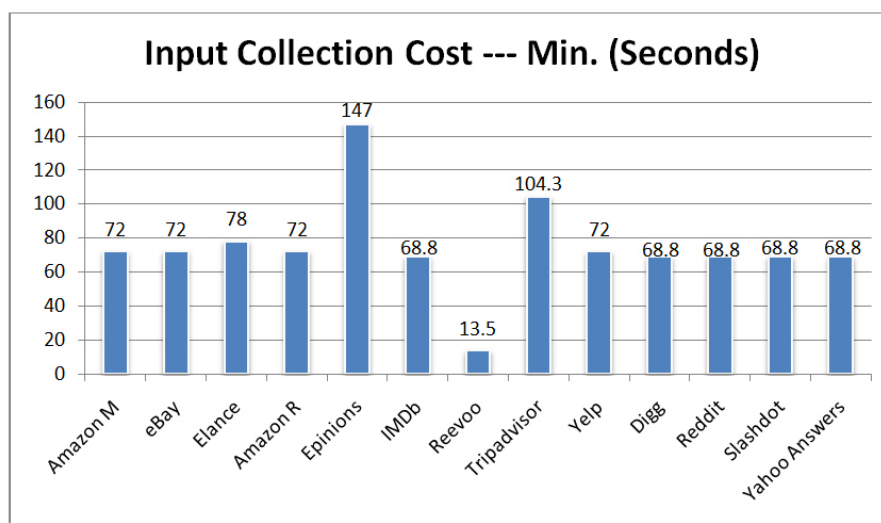
| Sites | Minimum Collection Setting | Maximum Collection Setting (2000 words) | 150-word Setting (150 words) |
|-------------|------------------------------|--|---|
| Amazon M | | 400 characters | 400 characters |
| eBay | | 80 characters | 80 characters |
| Elance | | 2000 characters | |
| Epinions | 20 words for the full review | 12 characters for one short review, 15 words for the other 3 short reviews, 2000 words for the full review | 12 characters for one short review, 5 words for the other three short reviews and 150 words for the full review |
| IMDb | | 1000 words | |
| Tripadvisor | 50 characters | | |

Table 6.2: Sites with Special Limitations

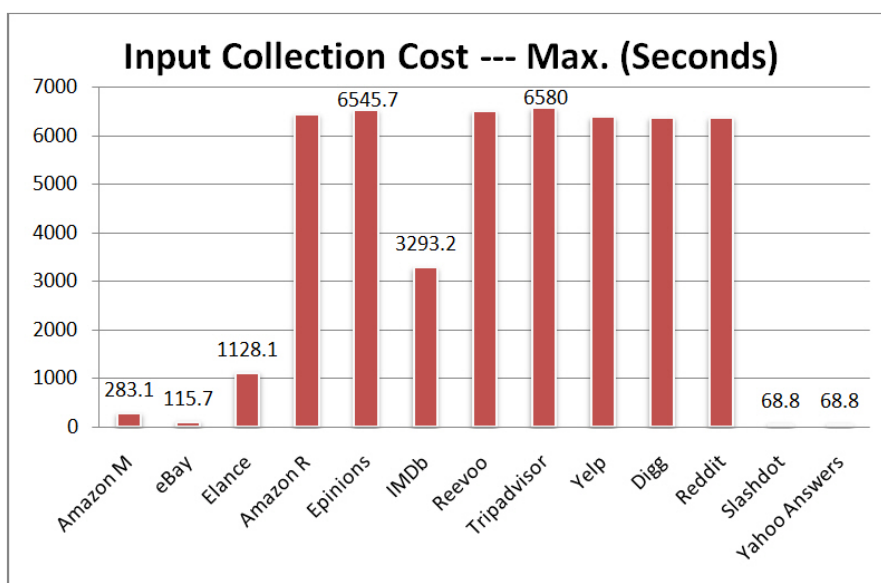
Sometimes reputation systems may ask for some ‘short reviews’, which are review titles or a few words on the pros and cons of the targets. The minimum and 150-word settings take 5 words for short reviews and the maximum setting uses 20 words. Moreover, some systems specify their words limitation clearly, for example they may allow no more than 1000 words for a full review, in this case, all the settings will follow their own requirements if there is a conflict.

Table 6.2 shows sites with special limitations. It can be seen that all C2C systems have maximum words limitations of reviews. IMDb is the only other site that has maximum limitation, whereas Epinions and Tripadvisor have the minimum words limitations.

Figure 6.5 compares the minimum and maximum settings of the collection costs. The minimum collection costs range from 13 seconds to 150 seconds, whereas maximum costs are between 60 seconds and over 6000 seconds. Under the minimum setting, most sites only collect one rating and one text review, which makes browsing the pages (T_{br} in Equation (4.3), Section 4.2.4) the most time consuming action. Reevoo, which uses C_{1b} channel, thus has the lowest minimum collection costs. Within the other sites, IMDb, Digg, Reddit, Slashdot and Yahoo! Answers have the



(a) Minimum Setting



(b) Maximum Setting

Figure 6.5: Input Collection Costs (Min. and Max. Settings)

lowest costs because evaluators are allowed to leave only one rating. Not surprisingly, Epinions and Tripadvisor, which have the minimum words limitation have the highest minimum costs.

The only information Slashdot and Yahoo! Answers collect is a rating, which makes their minimum and maximum collection costs both fixed at 68.8 seconds (Figure 6.5b). Due to the maximum words limitations, all the C2C systems are also in the bottom end of the maximum chart. IMDb, which has a 1000 words maximum

limitation has a relatively lower maximum costs. All the other sites have a similar maximum cost which is more than 6000 seconds (around 100 minutes).

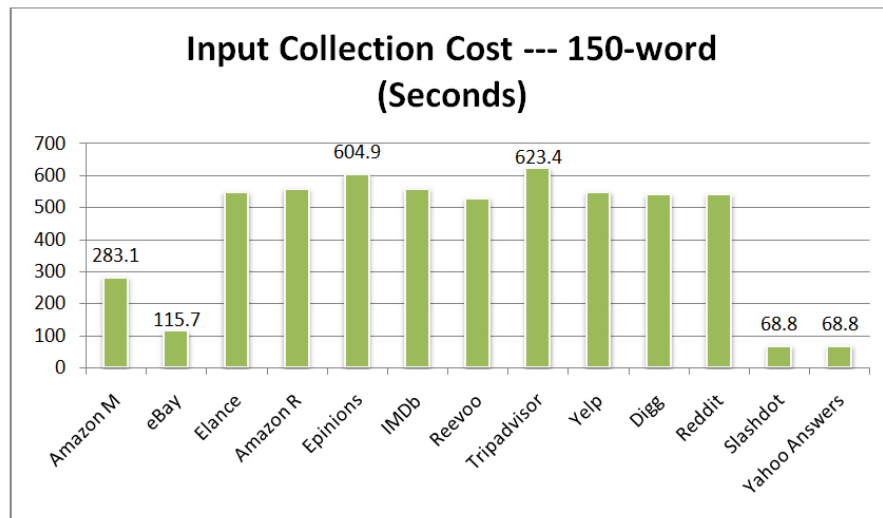


Figure 6.6: Input Collection Costs with 150-words

Figure 6.6 depicts the collection costs for the reputation information with the 150-word setting. It can be seen that Amazon M, eBay, Slashdot and Yahoo! Answers still have the lowest costs and all the other sites require 700-800 seconds collection time.

The three settings of the collection costs show that all the C2C systems have lower collection costs than the other sites due to their maximum words limitations. Yahoo! Answers and Slashdot also have low costs, whereas the other sites have relatively higher costs.

The collection costs have an influence on the proportion of evaluators who actually leave ratings and reviews (p_e , Section 4.2.2). Evaluators are more likely to leave ratings as it requires less time. For example, IMDb allows evaluators to leave ratings and reviews separately and as shown in Figure 6.7, the same movie receives much more ratings (16,792) than text reviews (211). Digg and Reddit have a similar policy and receive more ratings than text reviews as well.



Figure 6.7: The Numbers of Ratings and Text Reviews of the Same Target

6.3.2 Processing

6.3.2.1 Algorithms ($P1$, $P2$ and $P3$)

Section 4.3.1.1 has described several algorithms that are commonly used by commercial systems, for example, summation (SUM), mean (AVG) and percentage (PCT). Table 6.3 depicts the algorithms that reputation systems used for aggregating the target ratings, evaluator credibility (EC) and feedback.

| | Target Rating Algorithms ($P1$) | EC Algorithms ($P2$) | Feedback Aggregation Algorithms ($P3$) |
|-----------------|---------------------------------------|------------------------|--|
| Amazon M | AVG, PCT | | |
| eBay | SUM, PCT, AVG | SUM, PCT | |
| Elanca | Weighted Average for r_i ; AVG, PCT | | |
| Amazon R | AVG | SUM, PCT | SUM |
| Bizrate | AVG | | |
| Epinions | AVG | SUM | AVG |
| Google Shopping | AVG | | SUM |
| IMDb | AVG, WA, Median | | SUM |
| Reevo | AVG | | SUM |
| Tripadvisor | AVG | | SUM |
| Yelp | AVG | SUM | SUM |
| Digg | SUM | | SUM |
| Reddit | SUM, PCT | SUM | SUM |
| Slashdot | SUM | SUM | AVG |
| Yahoo! Answers | SUM | | |

Table 6.3: $P1$, $P2$ and $P3$ Algorithms

Target Rating Algorithms

E lance is the only system that uses multi-dimension overall ratings, i.e., it collects six ratings from evaluators and then aggregates the ratings with different weights to a final r_i :

$$r_{i,elance} = s_1 * 30\% + s_2 * 20\% + s_3 * 15\% + s_4 * 15\% + s_5 * 10\% + s_6 * 10\%$$

$s_1, s_2, s_3, s_4, s_5, s_6$ represent the ratings for ‘quality of work’, ‘responsiveness’, ‘professionalism’, ‘subject matter expertise’, ‘adherence to schedule’ and ‘adherence to cost’⁵.

For target aggregation algorithms, most systems prefer to use the *AVG* algorithm. However eBay and all the online communities use *SUM* for aggregation, which is because they all use the binary rating (‘-1, 0, +1’). Furthermore, all C2C systems and Reddit also provide a proportion of the ‘positive’ ratings, which use the *PCT* algorithm (Section 4.3.1.1). Elance and Amazon M also use *PCT* to generate the proportion of positive ratings, though they collect ratings from 1 to 5 and define ‘1,2’ as negative ratings, ‘3’ as neutral and ‘4,5’ are the positive ones.

IMDb provides two more overall ratings, one is the weighted average and the other is the median. The weighted average is introduced in order to ‘eliminate and reduce attempts made by individuals more interested in changing the current rating of a movie than giving their true opinion of it’⁶. The median is the number separating the lower half from the higher half of a set of ratings.

EC Algorithms

Systems that have evaluator credibility mechanisms all choose to use *SUM* to calculate the credibility scores. eBay and Amazon R also use *PCT* to estimate the positive ratings the evaluators have received.

Feedback Aggregation Algorithms

⁵Elance (2011). ‘Leaving feedback for a provider’. <http://help.elance.com/forums/30970/entries/34628>; Last Accessed 15 January 2011.

⁶IMDb (2011). ‘Weighted average ratings.’ http://www.imdb.com/ratings_explained; Last Accessed 15 January 2011.

All the C2C systems and Bizrate only collect text feedback, thus they do not use any feedback aggregation algorithms. Among the others, the most popular feedback algorithm is the *SUM* as well. That is because most systems collect binary feedback votes. Epinions and Slashdot use the average algorithm (*AVG*) as they use a Likert scale to collect ratings as feedback.

6.3.2.2 P4 Algorithm Robustness

As discussed in Section 4.3.2.1, the algorithm robustness can be estimated by the *breakdown point* (ϵ).

$$\epsilon_{sum} = \epsilon_{avg} = \epsilon_{pct} = \frac{1}{n} \text{ (} n \text{ is the number of reviews)}$$

According to Table 6.3, most systems only use *SUM*, *AVG* and *PCT*; thus their robustness are $\frac{1}{n}$. IMDb, however, also uses the median and a self-defined weighted average algorithm to present the rating results. The robustness of median is (Garcin et al., 2009):

$$\epsilon_{median} = \begin{cases} \frac{1}{2} + \frac{1}{2n} & \text{if } n \text{ is odd} \\ \frac{1}{2} & \text{if } n \text{ is even} \end{cases}$$

Because the site keeps the algorithm for calculating the weighted average private, its robustness cannot be estimated.

In general, most systems has a similar robustness, while IMDb has a better robustness on one of its target aggregation algorithms.

6.3.2.3 P5 Update Frequency and P6 Algorithm Complexity

As discussed in Section 6.2.2, there is no discussion in these two criteria because most systems do not disclose relevant information to the public.

6.3.2.4 P7 System Complexity

The system complexity refers to the features reputation systems provide (Section 4.3.2.3).

This research uses Equation (6.1) to estimate the system complexity:

$$\text{System Complexity} = N_{rdd} + N_{eca} + N_{fda} + N_{tm} + N_{isf} \quad (6.1)$$

- N_{rdd} refers to the number of descriptive dimensions that are used to illustrate aggregated information. Section 6.3.3.4 discusses the details on the assessment of the descriptive dimensions.
- N_{eca} and N_{fda} are the number of evaluator credibility algorithms and feedback aggregation algorithms that the systems use (Section 6.3.2.1).
- N_{isf} is the number of information sorting/filtering (ISF) dimensions that a system provides. Section 6.3.3.5 discusses the details of ISF dimensions. It should be noted that some systems can sort and filter by the same dimension, in this case, it is calculated as one feature rather than two.
- Finally, N_{tm} is the number of extra aggregated ratings that reputation systems provide for different time periods, which is discussed in Section 6.3.3.3. For example, if a system presents overall ratings for recent 3-month, 6-month and lifetime overall ratings, then $N_{tm} = 2$.

It should be noted that some features may require more system resources. Because most of the algorithms and features provided are simple. This research assumes that all features require the same resources. In other words, the systems complexity can be compared by the number of total features the system provides.

Figure 6.8 compares the system complexities. It shows that most sites are able to provide 6-10 features. However, IMDb, which is the most complicated system, provides 18 features. It illustrates 11 descriptive dimensions and provides 6 ISF dimensions, both are at the top of all sites.

On the contrary, Yahoo! Answers only provides 1 feature — the overall rating. In other words, it does not have any evaluator credibility mechanism, feedback loop mechanism nor ISF functions. In general online communities are slightly less complicate than the other types of system.

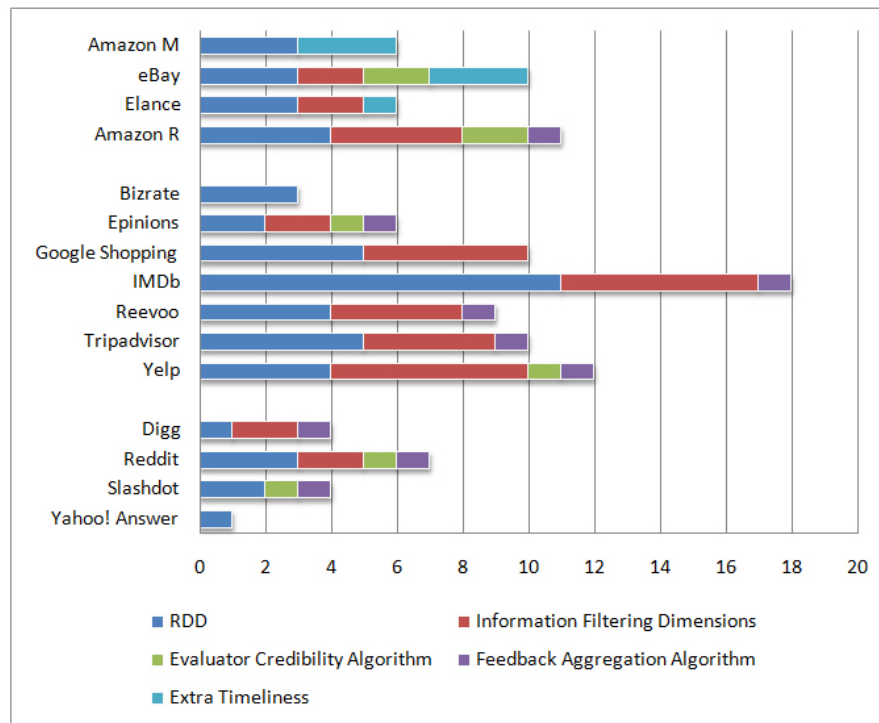


Figure 6.8: System Complexity Comparison

6.3.3 Output

6.3.3.1 O1 The Set of End Users

All the sites allow all their visitors to obtain the reputation information. However some systems reserve more information for specific users.



Figure 6.9: Different Information Presentation from Yahoo! Answers

Yahoo! Answers shows different information for general site visitors and ‘signed in’ users. Figure 6.9a shows that before signing into the site, users can only find the number of positive ratings, i.e., how many people have voted the answer as ‘good’;

while after signing in, users can see the number of both positive and negative ratings. Similarly, Epinions hides the feedback information for general visitors. Users need to sign in to the site before they can find out what other users think of the reviews. IMDb provides the Top 250 Movie ranking for all the Internet users but only allows IMDbPro users, which have paid a subscription fee, to access the Top 500 Movie rankings.

6.3.3.2 O2 Dissemination Method

All systems can disseminate information through their websites, while some sites provide more methods:

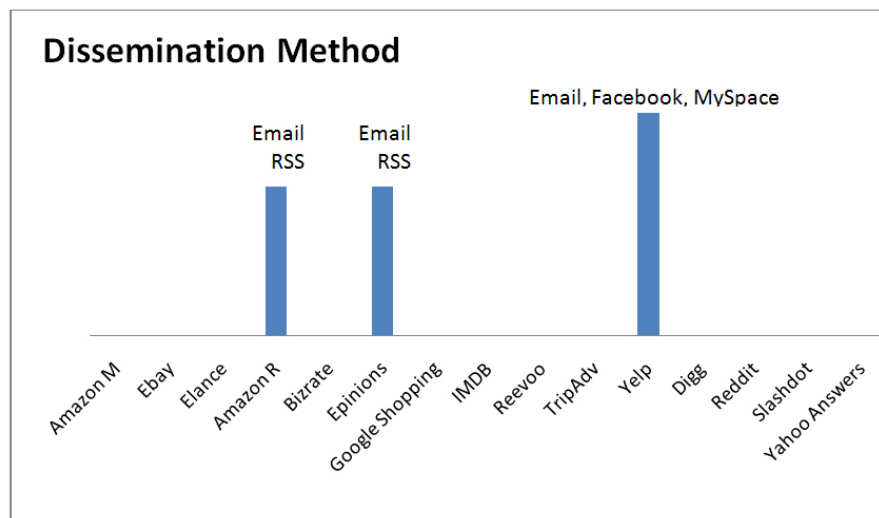


Figure 6.10: Dissemination Methods

- Amazon R builds an RSS feed for each reviewer, including all the reviews they have written. End users then can subscribe to any reviewer and get the latest review via the RSS. Amazon R can also send emails to the end user if a comment (considered as feedback in the research) has been left for the specified review.
- Epinions also supports both email alerts and RSS. The email alerts function sends email to the end users when a new review is left for the specified target. The RSS function is very similar to Amazon R's, which allows end users to subscribe to specified reviewer's review list.

- Yelp allows users to email reviews to their friends. Furthermore, the site also allows users to share reviews through Social Networking Sites (SNS), such as Facebook and MySpace.

Figure 6.10 shows that in general, Amazon R, Epinions and Yelp have more dissemination methods other than web pages publishing. The other sites do not provide any other dissemination methods.

6.3.3.3 O3 Timeliness

Figure 6.11 illustrates the distribution of timeliness selections. It shows that all C2C systems choose to present both short term and life time aggregations, and all the other sites, except Bizrate, only show the life time aggregation. Bizrate only presents information on the most recent 90 days.

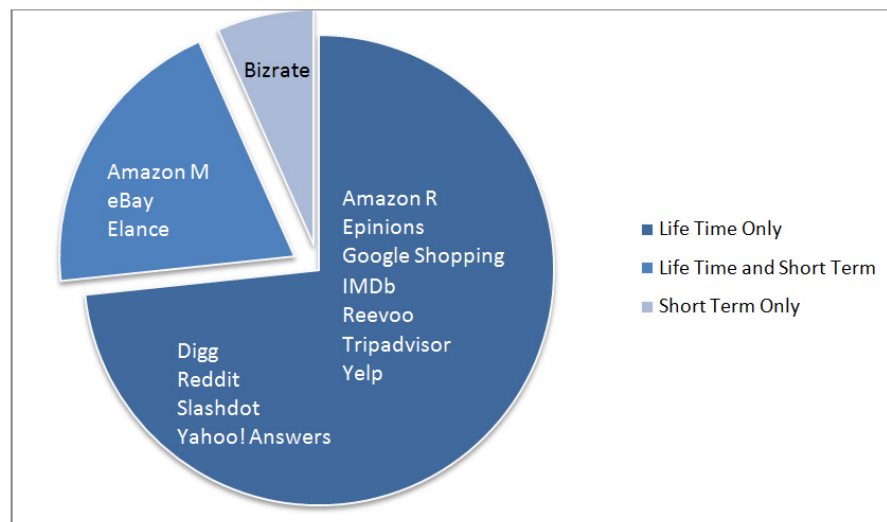


Figure 6.11: Timeliness Distribution

One possible reason that why C2C systems show both life time and short term information is that the targets of these systems are services, which are much easier to change their quality than other targets. For example, a seller may improve their service quality after receiving negative reviews, while a book cannot change its quality no matter what reviews it receives. Therefore, with the various time period aggregations, C2C systems can keep sensitive to the possible changing of quality.

6.3.3.4 O4 Descriptive Dimensions

This criterion estimates how many dimensions reputation systems use to illustrate the aggregated information. This research estimates this criterion with the number of the dimensions rather than the number of dimension ‘values’. For example, if two sites both present the percentage of individual rating score. One of them accepts ratings from 1 to 5 and the other accepts 1 to 10. Therefore, the former has five values to calculate and present, whereas the latter has ten. In this case, both sites are counted to have one dimension.

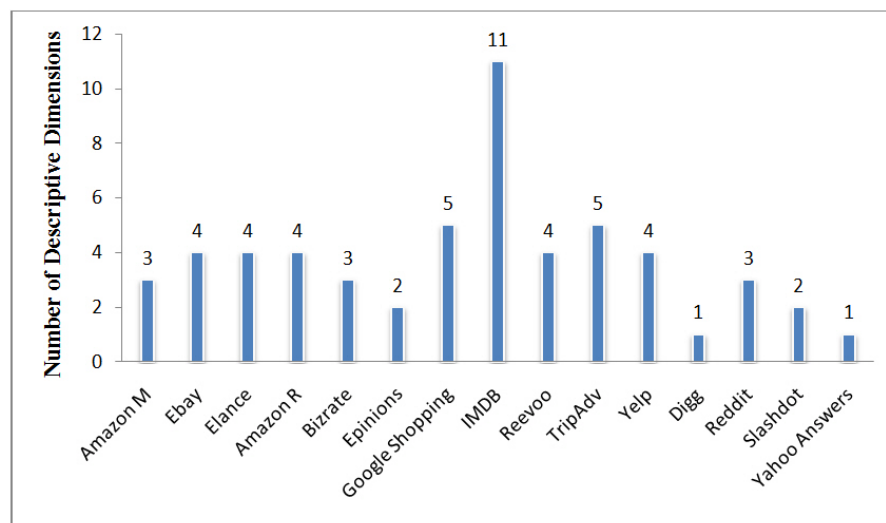


Figure 6.12: Descriptive Dimensions

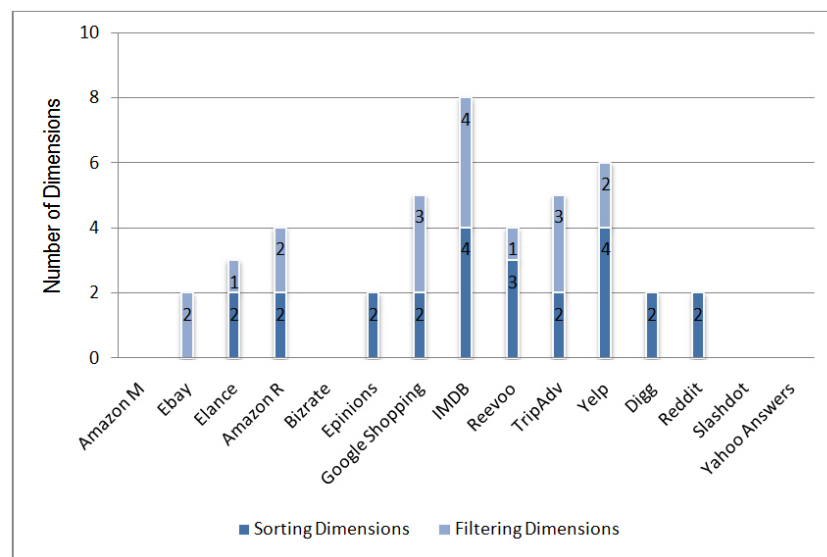
Among the 15 sites, IMDb uses more dimensions to describe the aggregated information than the others (Figure 6.12). That is because it provides three central tendency calculations, mean, weighted mean and median and illustrates a detailed demographic breakdowns, which presents detailed rating distribution of people in different gender, age and their locations. Digg and Yahoo! Answers, which present the least dimensions, only show an overall rating result.

In addition to the basic overall rating, the three C2C systems and Reddit also show the percentages of positive ratings. All the review centers and C2C systems provide the number of total ratings and reviews. Digg and Reddit can show the number of text comments but do not present the number of total ratings. Reddit, Slashdot and most review centers, except Bizrate, Epinions and Reevoo, depict the

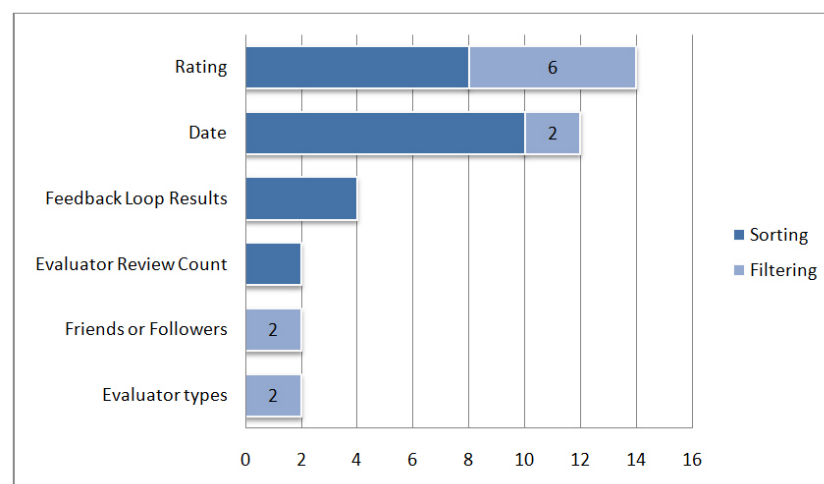
detailed rating distributions. One special feature of Reevoo is that the site shows the ratings by different type of reviewer.

Text review aggregation is not as popular as ratings; only four sites provide aggregated reviews. Amazon R compares the most ‘helpful’ positive review with the most ‘helpful’ negative review. Google Shopping, Tripadvisor and Yelp automatically generate review highlights.

6.3.3.5 O5 Information Filtering



(a) Sites Comparison



(b) Dimensions Comparison

Figure 6.13: Information Sorting and Filtering Comparison

The performance of information filtering is assessed by how many dimensions each site has to sort or filter the individual information. As shown in Figure 6.13a, Amazon M, Bizrate, Slashdot and Yahoo! Answers do not have any filtering/sorting function. The other sites can filter or sort reviews by at least two dimensions: the submitted dates and the ratings.

IMDb, once again, has the best performance within the 15 sites. IMDb users can both sort and filter by rating and submitted dates. They can also sort reviews by the feedback loop results (the number of helpful votes) and the number of reviews the reviewer has left. Furthermore, the site allows users to filter reviews by the characteristics of the reviewers, such as, age and country.

Similarly, Reevo and Tripadvisor also enable users to filter reviews by the different types of reviewer. As Google Shopping collects reviews from other reputation systems, the site can filter reviews based on the source sites.

Some sites started to add SNS features, for example, users of Tripadvisor and Yelp can add others as *friends* or *followers*, whose reviews can then be filtered. Amazon R and Yelp are the only two sites that allow end users to search within reviews. Such a function is very useful for users if they want to explore a specific factor of the target.

In general, the most popular dimensions are ratings and date (Figure 6.13b). All the sites have the sorting or/and filtering function can do so both by ratings and the date. The helpfulness of the review is another popular dimension, which is generated by the results of the Feedback Loop. Some sites also allow users to sort reviews by the evaluators characters, such as evaluators' type and the number of reviews the evaluator has left to other targets. In general, with the growing number of reviews, it seems reputation systems need to enhance their information filtering function by adding more dimensions.

6.3.3.6 O6 Evaluator Information

Evaluators are important for reputation systems. By providing more information about the evaluators, end users can make a better judgment on their reviews. Slashdot and Yahoo! Answers are the only sites that do not provide any information of

the evaluators, not even the user IDs (Table 6.4). Amazon M only displays the user IDs of their evaluators.

| | Review History | User ID | Credibility | Feedback Results | Others |
|-----------------|----------------|---------|-------------|------------------|-------------------|
| Amazon M | | ✓ | | | |
| eBay | ✓ | ✓ | ✓ | | |
| Elanca | ✓ | ✓ | | | |
| Amazon R | ✓ | ✓ | ✓ | ✓ | Real Name |
| Bizrate | | ✓ | | | |
| Epinions | ✓ | ✓ | ✓ | | |
| Google Shopping | | ✓ | | | |
| IMDb | ✓ | ✓ | | | |
| Reevo | | ✓ | | | Self-defined Type |
| Tripadvisor | ✓ | ✓ | | | |
| Yelp | ✓ | ✓ | ✓ | | No. of Friends |
| Digg | ✓ | ✓ | | | No. of Friends |
| Reddit | ✓ | ✓ | ✓ | ✓ | |
| Slashdot | | | | | |
| Yahoo! Answers | | | | | |

Table 6.4: Evaluator Information

Bizrate and Google Shopping collect reviews from other sites, such as, online shops or other reputation systems. Limited by the way of information collection, they are not able to present much information other than the user IDs. Although, Reevo also depends on other online shops to collect reviews, it requires evaluators to define their own types when leaving reviews. For example, when leaving reviews for digital cameras, evaluators can define themselves as ‘Keen amateur’, ‘Experienced amateur’, ‘Point&Shoot’, ‘Professional photographer’ and ‘other’ types. Thus, the site can show the self-defined types of evaluator.

All other sites can show evaluators’ review histories, i.e., all the reviews they have submitted. eBay, Amazon R, Epinions, Yelp and Reddit have evaluators’

credibility mechanisms. Thus all these sites show the credibility of the evaluators. Amazon R and Reddit also present the number of positive feedback the evaluators have received. Yelp and Digg show the number of friends each evaluator has.

6.3.3.7 O7 Feedback Information

This criterion measures how reputation systems present feedback results. Yahoo! Answers is the only site that does not have feedback mechanism. Although Slashdot can collect feedback, it does not display the results to the public. Therefore, neither Slashdot nor Yahoo! Answers shows any feedback information (Table 6.5).

| | Feedback Provider | Text Feedback | Positive Votes | Negative Votes | Total Votes | Final Results |
|-----------------|-------------------|---------------|----------------|----------------|-------------|---------------|
| Amazon M | ✓ | ✓ | | | | |
| eBay | ✓ | ✓ | | | | |
| Elance | ✓ | ✓ | | | | |
| Amazon R | T^* | ✓ | ✓ | | ✓ | |
| Bizrate | ✓ | ✓ | | | | |
| Epinions | ✓ | ✓ | | | | ✓ |
| Google Shopping | | | ✓ | | ✓ | |
| IMDb | | | ✓ | | ✓ | |
| Reevo | | | ✓ | | ✓ | |
| Tripadvisor | T^* | ✓ | ✓ | | | |
| Yelp | T^* | ✓ | | | | ✓ |
| Digg | T^* | ✓ | ✓ | ✓ | | ✓ |
| Reddit | T^* | ✓ | | | | ✓ |
| Slashdot | | | | | | |
| Yahoo! Answers | | | | | | |

T^* indicates the system only shows text feedback providers.

Table 6.5: Feedback Loop Information

Bizrate and the three C2C systems only collect text feedback. The C2C systems allow the targets, i.e., sellers, buyers and freelancers, to provide information from

their perspective. Similarly Bizrate allows their partner shops to leave responses to customers' reviews. All the systems that collect text feedback also show the information of feedback providers. Epinions is the only site that can identify the provider of ratings.

Amazon R, Google Shopping, IMDb, Reevoo, Digg and Reddit collect binary feedback ratings, i.e., positive and negative votes. As shown in Table 6.5, most of them present the results by the number of positive votes and number of total votes, such as, '40 of 43 people found the review helpful'. Reddit only shows the '*final results*' of the voting, i.e., the number of positive votes minus the number of negative votes. Digg, however, presents both numbers of positive and negative votes as well as the final results. Yelp and Tripadvisor not only provide end users' feedback but also show the comments from the 'owners of the shops' (their targets).

Epinions use a Likert rating scale to collect feedback ratings, by which users can choose from five levels: 'off topic', 'not helpful', 'somewhat helpful', 'helpful' and 'very helpful'. Yelp, however, chooses to use 'tag'-format feedback. The site asks users to label reviews as 'Useful', 'Funny' or 'Cool'.

6.3.4 Feedback Loop

6.3.4.1 F1 Feedback Loop Function

As discussed in Section 4.5.1, the Feedback Loop has three main functions. Function 1 uses feedback to identify the helpfulness of the reviews, Function 2 uses feedback to alert the non-review spams and Function 3 uses feedback to provide more information.

Among the 15 sites, Yahoo! Answers is the only one that does not have any feedback loop. Figure 6.14 illustrates how the other sites utilize feedback loops. 7 sites rely on feedback for more than one functions. Within them, IMDb is the only one that use feedback for Function 1 and 2, the others use feedback for all the three functions. Bizrate and the C2C systems only use the feedback loop to provide further information (Function 1), whereas Amazon R, Epinions, Digg and Reddit adopt feedback loop for all the functions.

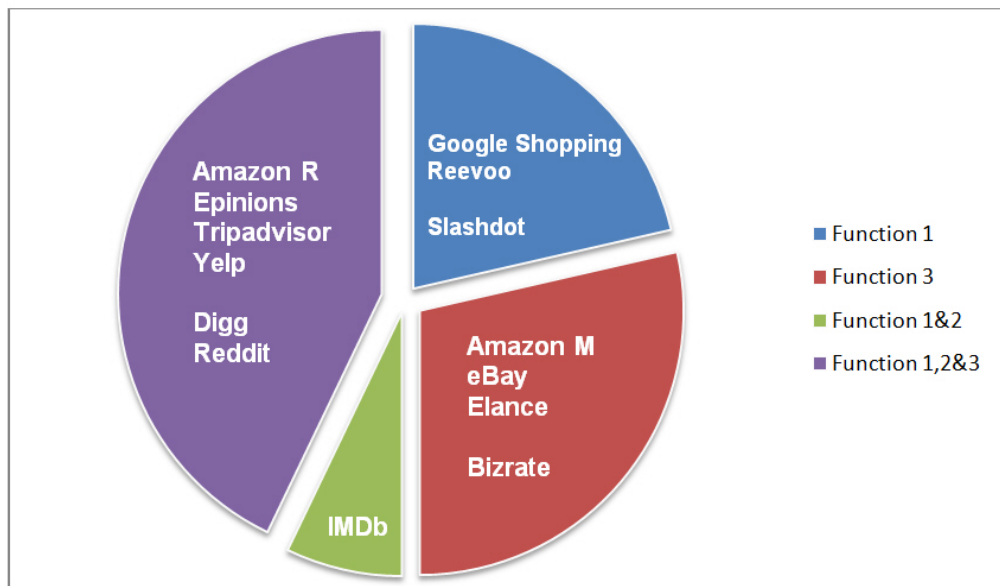


Figure 6.14: Feedback Function

6.3.4.2 F_2 The Set of Feedback Providers

The set of feedback providers (U_{fe}) indicates the number of feedback providers that leave feedback to the reviews. According to Section 4.5.2, $|U_{fe}| = |U_{fq}| * p_{fe}$. Similar to the assessment of I_2 (the set of evaluators Section 6.3.1.2), this section also assumes all systems have the same p_{fe} .

As discussed in the previous section, Tripadvisor and Yelp allow both end users and targets to leave feedback, which means they have two sets of feedback providers. One only consists of 1 provider — the target (shop owners) and the other set including all the registered users. Figure 6.15 shows the bigger sets only.

Google Shopping, Reevo and Yelp allow people to give feedback without registering with the sites. In contrast, Bizrate and all the C2C systems, which adopt Feedback Loop for Function 1, only accept one evaluator, i.e., the target, to provide feedback.

Slashdot allows ‘the oldest 92.5% of accounts’ users to be feedback providers⁷. All the other sites allow all registered users to give feedback.

⁷Malda, R. C. (2011). ‘Slashdot faq’. <http://slashdot.org/faq/>; Last Accessed 15 January 2011.

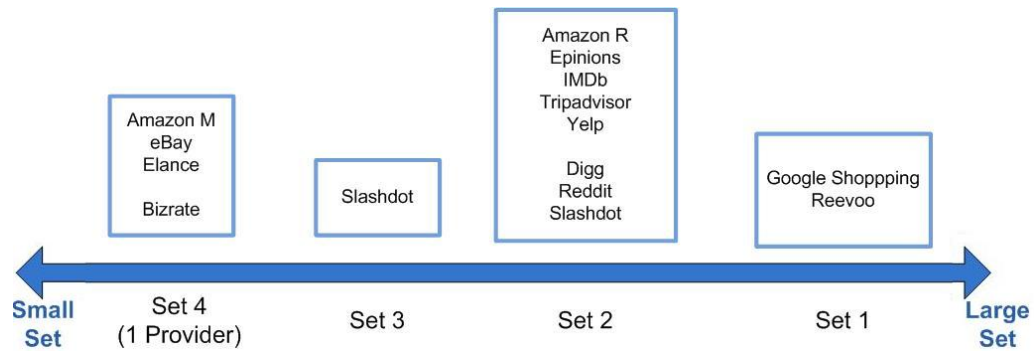


Figure 6.15: Sets of Feedback Providers

6.3.4.3 F3 Feedback Format and Breadth

This criterion specifies the feedback format and how much information has been collected. Because the results of Function 2 (the spam report) are not published to the users, the discussion in this section excludes the Function 2 related information.

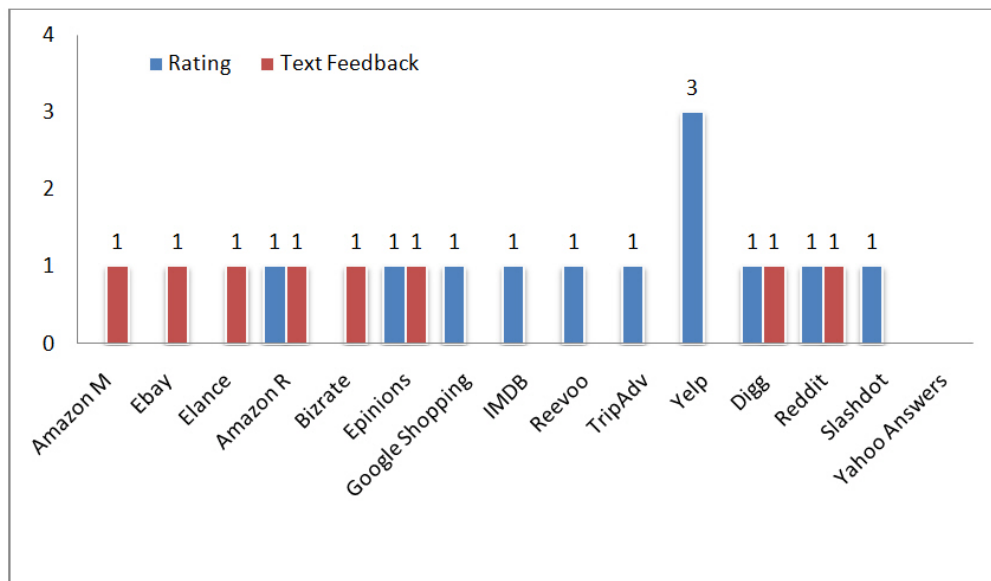


Figure 6.16: Feedback Breadth

According to Figure 6.16, which compares the formats and breadth, most systems only collect one rating and/or one text feedback. Yelp uses tag-like ratings, which allows feedback providers to rate the review as ‘Useful’, ‘Funny’ or ‘Cool’.

Comparing to the input information format and breadth (Section 6.3.1.5), reputation systems collect much less feedback information. Furthermore, unlike Input, reputation systems usually allow feedback providers to give ratings and text feedback

separately.

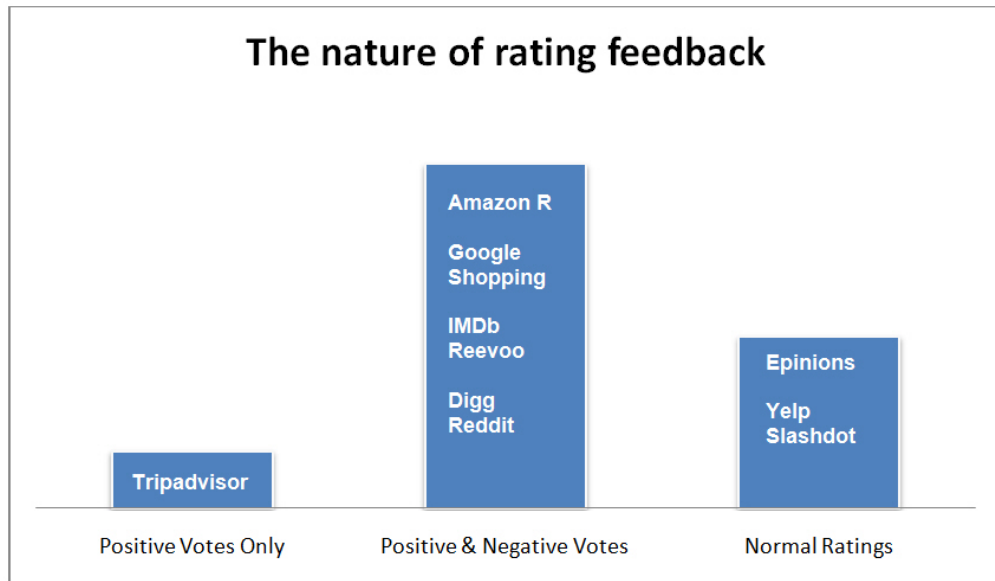


Figure 6.17: The Nature of Rating Feedback

Figure 6.17 shows the nature of rating feedback, i.e., the format of ratings. Tripadvisor is the only one that collects positive votes without negative ones, while other similar sites, Google Shopping, IMDb, Reevo, Amazon R, Digg and Reddit collect both positive and negative votes. Slashdot and Epinions allow feedback providers to select scores from a range of ratings, while Yelp uses the tag-like ratings.

6.3.4.4 $F4$ Feedback Loop Level

Within all 14 sites that have Feedback Loops, five sites have multiple levels of feedback (Figure 6.18): eBay, Amazon R, Epinions, Digg and Reddit. The aim of having multiple levels of feedback loop is to provide a community for users to discuss and communicate on specific topics. All the other sites only provide one level of feedback loop, which means a feedback provider can only leave one text comment or rating to the review.

6.3.4.5 $F5$ Feedback Loop Collection Cost

The feedback collection cost is calculated by Section 4.5.3. Similarly to the estimation of Input Collection Costs (Section 6.3.1.6), this section sets up two settings to

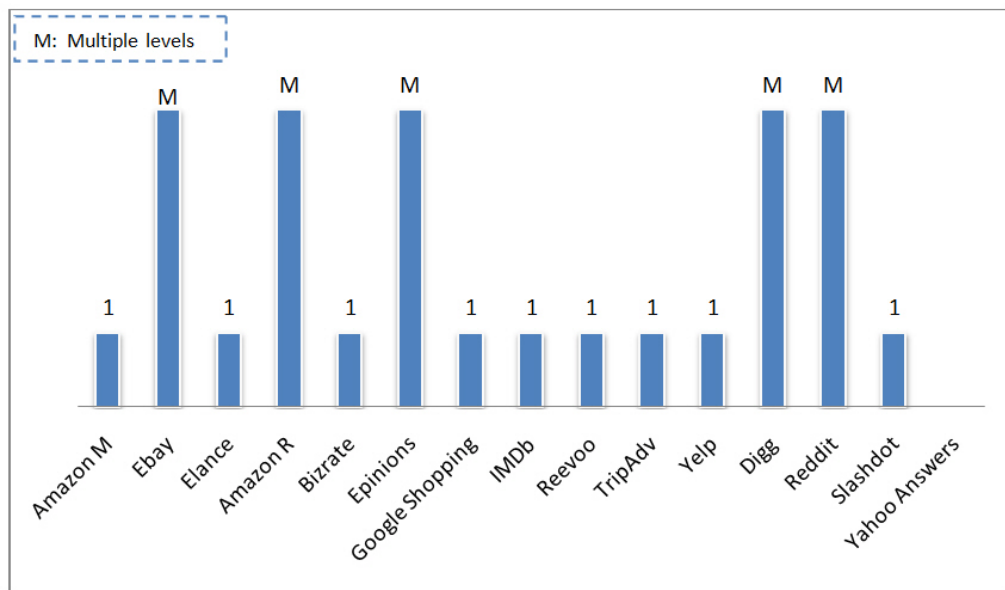


Figure 6.18: Feedback Level

compare the feedback collection costs:

- **Minimum costs** aims to compare the time it takes feedback providers to give least information. The setting assumes the feedback provider only need to provide the required information with the least cost. For example, if a feedback provider can leave either a rating or a text feedback, the setting assumes they only provide one ratings.
- **150-word setting** assumes feedback providers not only give ratings but also leave 150-word text feedback. People tend to write less words in feedback than in reviews; thus the setting chooses to use the 150 words, the average words in review, as the maximum setting for feedback.

Figure 6.19 shows the collection costs which are calculated by Equation (4.26) in Section 4.5.3 with the two settings. It can be seen that the collection costs of Google Shopping, IMDb, Reevoo, Tripadvisor and Slashdot are fixed at 68.8 seconds, because they only collect one rating as their feedback. Yelp allows feedback providers to make 1, 2 or 3 ratings, which means, the difference between the two settings is around 3 seconds (the data of Tripadvisor and Yelp only calculate the feedback provided by the end users). For all the other systems allowing text feedback, the costs of the 150-word setting are much higher.

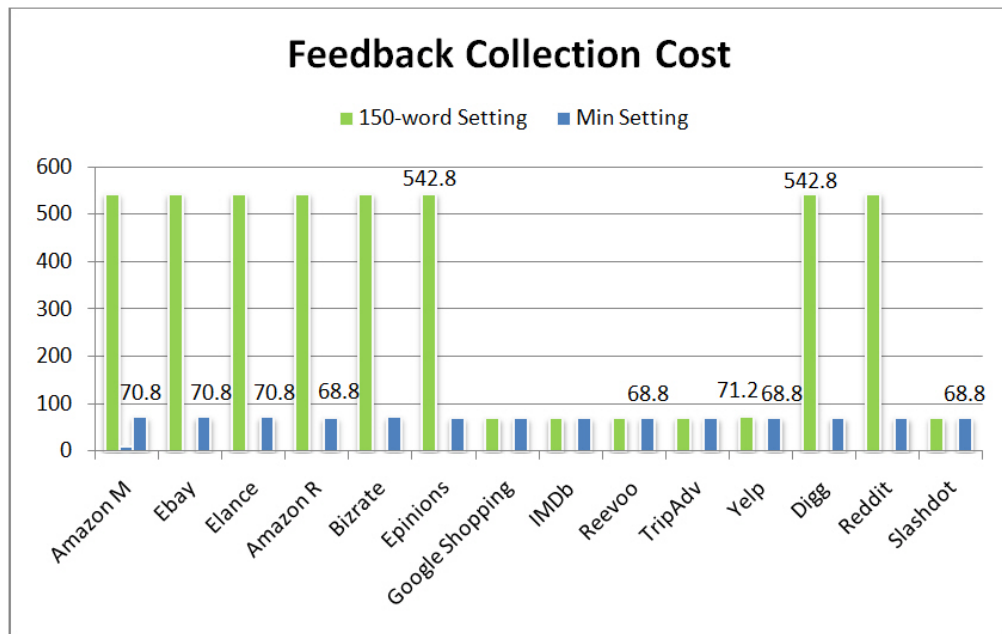


Figure 6.19: Feedback Collection Cost

6.3.5 Storage

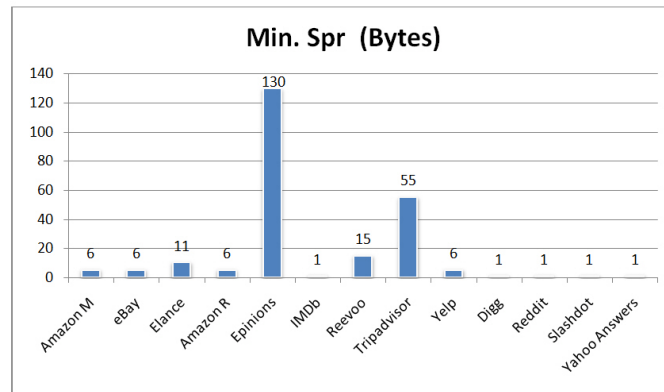
6.3.5.1 S1 Storage Costs of Input Information

Section 4.6.1 discussed the storage cost of input information and Equation (4.27) showed that the total storage cost of input information (S_{ip}) is related to the number of reviews (N_{tr}) and the size of the information an evaluator submits (S_{pr}). Because only two sites accept rich media information and very few evaluators provide this information, the discussion only focuses on the ratings and text reviews.

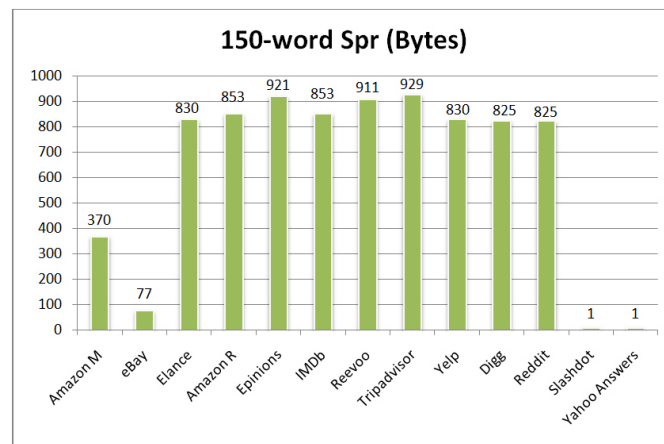
The size of S_{pr} is decided by the information format and breadth. Similar to the discussion of Input Collection Costs (Section 6.3.1.6), this section also compares the S_{pr} by three settings: minimum setting, maximum setting and 150-word setting. The settings are as the same as the ones in Section 6.3.1.6.

Figure 6.20a, 6.20b and 6.20c compares the size of S_{pr} based on the three settings. Because it is not possible to retrieve the data of their input information, Bizrate and Google Shopping's storage costs are excluded.

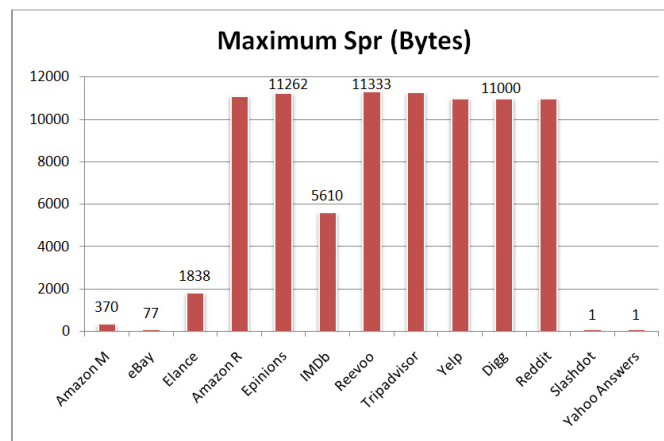
As expected, Slashdot and Yahoo! Answers have the lowest storage costs in all three settings because they only collect one rating. Amazon M and eBay also have relative low costs due to their maximum words limit. Epinions, Tripadvisor and



(a) Min. Setting



(b) 150-word Setting



(c) Max. Setting

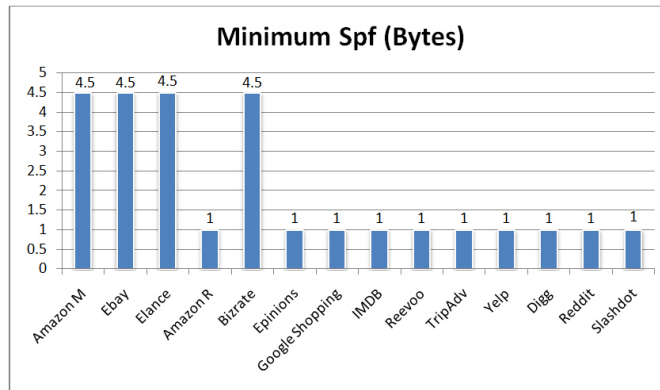
Figure 6.20: The Value of S_{pr} (Bytes)

Reevoo has the highest storage costs in all three settings. IMDb has a relative low maximum cost is because the site has a maximum review words limitation.

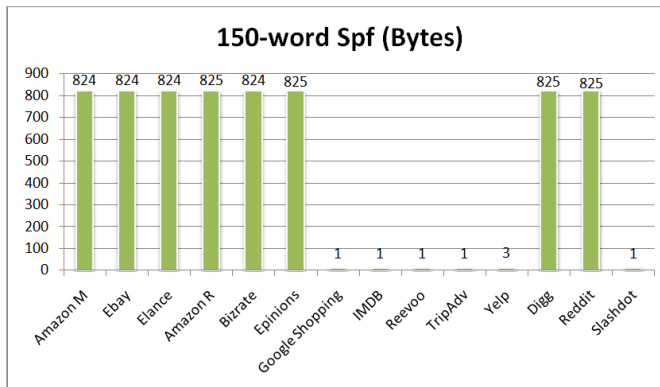
The value of N_{tr} is decided by $|U_e|$ (Section 4.2.2). The discussion in Section 6.3.1.2 showed that by assuming all the sites have the same U_v and p_e , Amazon M, eBay, Elance, Slashdot and Yahoo! Answers have the smallest U_e . Considering these sites also have relative low S_{pr} , they are then supposed to have lower value of S_{ip} than the other ones. Other sites have both larger S_{pr} and U_e , therefore have a higher input storage costs.

6.3.5.2 S2 Storage Costs of Feedback

Feedback storage cost is related to the level of the feedback loop (L), the number of feedback provider (U_{fe}) and the size of one feedback (S_{pf}) (details are discussed in Section 4.6.2). Yahoo! Answers is excluded from the following discussion because it does not have any feedback loop.



(a) Min. Setting



(b) 150-word Setting

Figure 6.21: The Value of S_{pr} (Bytes)

Figure 6.21 compares the value of S_{pf} based on the minimum and 150-words settings, which are same as ones in Section 6.3.4.5. It showed that the sites only collect ratings as feedback, Google Shopping, IMDb, Reevo, Tripadvisor, Yelp and Slashdot, have the similar S_{pf} for both settings (again, the data of Tripadvisor and Yelp only calculate the feedback provided by the end users). Amazon M, eBay, Elance and Bizrate only collect text feedback, thus they have relative high S_{pf} on both settings.

As discussed in Section 6.3.4.2, Google Shopping, Reevo and Yelp have the biggest set of feedback providers. However, these sites only collect ratings, which occupy much less storage than the text feedback. For example, the storage size of 100 ratings, which are made by 100 providers, is 100 Bytes, whereas the size of one 20-words feedback is 109 bytes. In other words, a system with a larger U_{fe} does not guarantee a higher S_{fd} .

The level of the feedback loop also has a great influence on the storage. Amazon R, Epinions, Digg and Reddit accept multiple levels and their S_{fd} are relatively higher than the single level systems. Although eBay also allows multiple level feedback, the system only allows one provider, which means their feedback storage cost is supposed to be lower than the others.



Figure 6.22: The Comparison of S_{fd}

Figure 6.22 compares the size of S_{fd} . Amazon R, Epinions, Digg and Reddit have the highest feedback storage costs, because of their multiple levels, larger U_{fe}

and collection of both ratings and comments.

Within the sites that have the lowest storage costs, Amazon M, eBay, Elance and Bizrate only allows one provider to leave text feedback, whereas the other sites allowing more evaluators but only accept ratings. As discussed above, it is difficult to distinguish the size of these two kinds of sites.

6.3.5.3 S3 Storage Costs of Processing Information

As discussed in Section 6.2.2, there is no evaluating this criterion because reputation systems do not disclose relevant information to the public.

6.4 Conclusion

The SERS model is designed to evaluate different types of reputation system in the same context. Section 4.7 grouped all criteria into three types: classification criteria, measurement criteria and cost criteria. The results in this section showed that the classification criteria can show the differences between reputation systems. For example, Figure 6.1 in Section 6.3.1.1 showed that most systems use C_{1a} channel collect information, whereas Bizrate and Reevoov choose to use C_{1b} and Google Shopping is the only site prefer C_2 .

In addition, the measurement criteria are capable of comparing the performance of different sites. For instance, Section 6.3.3.4 compared the reputation systems by the descriptive dimensions. The results showed that within all sites, IMDb has the best performance and Digg and Yahoo! Answers are the worst.

The cost criteria can identify the costs of the sites. For example, results in Section 6.3.1.6 and Section 6.3.5.1 showed that Epinions has the highest costs for both input collection and storage. Furthermore, according to Section 6.3.2.4, IMDb has the most complicated reputation system and Yahoo! Answers has the simplest one.

Moreover with the results of this section also showed the drawbacks of reputation systems:

- A lack of adequate information about evaluators.

Most systems cannot identify the granularity of evaluators. At the moment, only 6 sites can identify the granularity and Epinions is the only one that is able to show the expertise level of the evaluators. A higher granularity can help end users estimate the quality of the reviews.

Moreover, though most sites can provide some information on the evaluators, evaluators' type and credibility are the main missing factors. Reevoo is the only site that shows the different evaluators types. Each consumer has different needs for the same target, thus reputation systems should differentiate their evaluators so that users can find the desired information. Furthermore, only five sites can identify the credibility of evaluators and present it to the users. By providing more credibility information of the evaluators, reputation systems can increase the trustworthiness of the reviews.

- A lack of reputation information format, multiple measurements and multiple quality factor ratings.

Currently, only two sites accept rich media reputation information, all the other sites only collect ratings and text reviews. One possible reason is that collecting rich media formats information may largely increase the costs of the system. First, it takes much more time for evaluators to make photos or videos than giving ratings or writing reviews, which results a higher input collection cost. Second, rich media information also requires more processing and storage costs.

At the moment, most sites use the 5- or 10-point rating scales and almost all of them only present the end users with the arithmetic mean of the overall rating results. However, the arithmetic mean only tells one aspect of the data set. To provide more details on the central tendency of a set of data, additional measurements should be calculated, such as, median or mode.

In addition, most sites also ignore the multi-dimensional nature of reputation. One overall rating is not sufficient to illustrate the quality of a target. For example, when searching for a camera, people want to know about many of its properties, including quality, ease of use, image quality and so on. Therefore,

reputation systems should present rating information on these aspects. eBay, Elance, Bizrate and Reevoo are the only four sites that present detailed multi-dimensional ratings.

- A lack of information filtering functions.

The information filtering function is still required for some sites. At the moment, four sites do not have any filtering function at all. Other sites need to add more dimensions for users. A search function, which can provide flexibility for users to find desired information, is also missing in most sites.

- More dissemination methods are in need.

Currently, most sites can only provide information via their websites. With more dissemination methods, reputation systems can spread information more effectively. As discussed in Section 4.4.1, RSS and email alters can let end users get the information more conveniently. Moreover, in recent years, SNS have changed the way people communicate with each other on the Internet. Some sites have integrated their services with Facebook, Twitter or other similar SNS. However at the moment, only Yelp allows users to share reviews with SNS. Enabling this feature can help reputation systems spread information more efficiently and attract new users.

As discussed in Section 2.3.4, reputation systems can be classified into three types: C2C systems, review centres and online communities. Based on the analysis of the assessment of 15 commercial sites, it can be found that C2C systems have similar performances on more criteria than the other two categories. In other words, C2C systems share more common characters than other categories.

In total, C2C systems have the same performances on 10 criteria, including *I1*, *I2*, *I3*, *P1*, *O2*, *O3*, *O7*, *F1*, *F2* and *F3*. Online communities have similar performances on 6 criteria: *I1*, *I3*, *I5*, *I6*, *P1* and *O3*. The only common character of review centres is *P1* (target rating algorithms), for which they all use *AVG* algorithm.

The results in this chapter indicated that the SERS model can be applied to

commercial sites and it can classify and measure them. The results also showed that both commercial web sites and academic researchers can benefit from the SERS.

Firstly, the above analysis has shown that there are four common problems with the current systems. Commercial sites may benefit from modifying or updating their systems accordingly. For example, reputation systems should collect more quality factors from evaluators and add more review filtering and sorting dimensions, so that end users can find their desired reviews more efficiently. The SERS can also help reputation systems to identify their own drawbacks. Different systems may have different needs in the designing of reputation systems. With the help of their own data, commercial sites can optimize their systems according to the SERS. In addition, newcomers can use the SERS as a guideline to design their systems.

Secondly, academic researchers can use the SERS to evaluate their own work. For example, if a new system is proposed, researchers can use the SERS to assess it or to compare it with other systems. Moreover, if more empirical data can be collected, researchers from different disciplines, such as computer scientists, mathematicians, economists and psychologists, can cooperate with each other on proposing an ideal system on the basis of the SERS.

In summary, the SERS can be used as a guideline for designing, optimizing and analyzing reputation systems from both industry and academic perspectives.

Chapter 7

Conclusion

7.1 Introduction

This thesis has aimed at building an evaluation model that can represent characteristics of reputation systems and assess different systems in the same context. As introduced in Chapter 1, reputation systems have been widely adopted by most e-commerce sites and other online companies. Research has shown that as long as a system can collect truthful information from sufficient reviewers, the aggregated information and individual reviews can effectively reflect the true quality of the target and then help other users to make decisions.

However, research to date has focused on measuring isolated systems. For example, much research concentrates on how C2C (consumer-to-consumer) marketplaces use reputation systems to build trust between strangers and the effectiveness of these systems (Section 2.4.1). Review centres also attracted a number of researchers, most of which analyzing whether the product review can influence the sales of the products (Section 2.4.2). Online communities, which utilize reputation systems to filter information, do not receive as much attention as the other types (Section 2.4.3).

Only a few studies aimed at comparing different systems together. However some of them only focused on decentralized systems and others concentrated on the rating algorithms, which is merely one aspect of reputation systems (Section 2.4.4).

The thesis fills in the gaps and aims at proposing a systematic evaluation model for assessing different systems.

7.2 Contributions

This research first systematically analyzed the entities and structure of reputation systems (Chapter 3). As discussed in Section 3.3, all reputation systems have five underlying components: Input, Processing, Output, Feedback Loop and Storage. Input collects ratings and reviews, which are then aggregated by the Processing. Output publishes the collected and processed information with meaningful forms. All the information is stored in the Storage. Some systems use Feedback Loops to control the quality of ratings and reviews.

The SERS (Systematic Evaluation of Reputation System) model, which was described in Chapter 4, defined a series of benchmark criteria for each component based on their characteristics. In total, 29 criteria have been defined and were grouped into: classification criteria, measurement criteria and cost criteria.

The analysis in Chapter 5 compared the criteria defined in SERS with the influential characteristics of reputation systems and measurement dimensions of Information Quality (IQ) and System Quality (SQ), which are the technical quality factors of Information Systems (IS). The results showed that the SERS not only can cover the major factors of reputation systems, but also can reflect most of the technical measurement dimensions of IS. Moreover, the SERS has defined a number of criteria that concentrated on the cost of reputation systems, which have been long ignored in other research.

The SERS has also been evaluated by applying it to 15 commercial sites that adopted reputation systems (Chapter 6). The 15 sites were selected based on their different types, roles and targets. The results discussed in Section 6.3 showed that the SERS can distinguish the characteristics of each site and assess both the performance and cost of them.

By analyzing the assessment of the sites, Section 6.4 identified a number of drawbacks of current reputation systems including the lack of information on the evaluators, less quality factors on targets, limited information filtering function and a lack of dissemination methods. Furthermore, the results also showed that C2C systems have similar performances on 10 out of 26 criteria, while review centres only share 1 criterion.

The SERS model can be used as a knowledge resource and guide for design and analyzing reputation systems for many different settings.

7.3 Criteria for Success

A number of criteria were specified in Section 1.2 as the judgment of the success of the research. This section discusses whether the research meets the criteria.

1. “A model that can represent the major characteristics of reputation system.”

As described in Section 3.3, the SERS was built based on the five underlying structure components of reputation systems. In addition, the criteria of the model were systematically defined according to each component characteristics.

Section 5.2 showed that the SERS has specified all the major characteristics that have been identified by the other research. Furthermore, the results in Section 5.3 showed that SERS can represent all the information quality dimensions and most of the system quality dimensions from Information Systems perspective. Although the SERS cannot reflect ‘response time’ and the technical aspect of system ‘reliability’, it is acceptable because the SERS focused on the intrinsic nature of reputation systems rather than the performance of the commercial sites.

2. “The model should consider the cost of reputation systems.”

As detailed in Section 4.7, in total, the SERS has defined seven cost criteria, including *I7* (Input Collection Cost), *P6* (Algorithm Complexity), *P7* (System Complexity), *F5* (Feedback Collection Cost), *S1* (Input Storage Cost), *S2* (Feedback Storage Cost) and *S3* (Processing Information Storage Cost). Although ‘money’ cost has not been considered in the research, the SERS can identify most of costs on information collection, processing and storage.

3. “The model can be empirically evaluated using samples taken from the commercial world.”

Chapter 6 chose 15 sites as examples and applied the SERS to evaluate them. The results in Section 6.3 showed that the SERS can assess all the sites with most

of the criteria. Due to the unavailable data of the sites, several criteria have not been discussed, including, *P5* (update frequency), *P6* (algorithm complexity) and *S3* (the storage cost of processing information). The results also have indicated a number of the drawbacks of these systems (Section 6.4).

4. “The model can compare and measure different types of system.”

The sites were selected in Chapter 6 include 3 C2C systems, 8 reviewer centres and 4 online communities. The targets of the sites covered services (transactions, jobs, hotels, local amusements, online shops), general products (books, electronic products), movies, stories and information (stories and news). The results in Chapter 6 showed that the SERS can compare and distinguish different systems.

7.4 Limitations

The proposed evaluation model (the SERS) focuses on the intrinsic nature of reputation systems; thus during the analysis and discussion, the factors of the web sites, with which reputation systems are integrated, have been ignored. However, practically these factors do have influences on the performance of systems. For example, a better web design can attract more site visitors and eventually, it can have impact on the number of evaluators. Therefore, future work can introduce web site evaluation criteria into the SERS and extend it to a more comprehensive version.

Although Chapter 6 has sampled 15 sites and collected data from them, in order to eliminate the influence of the web site, the assessment of some criteria has had to use theoretical data rather than actual data. For example, when assessing the set of evaluators (*I2*, Section 6.3.1.2), the best way to compare the number of evaluators would be comparing the number of reviews, which are left on each site, on the same target. However, the actual data is much influenced by the operation of the site, web design and other web site factors, which cannot reflect the true performance of the system.

Another limitation of this research is that criteria defined in the SERS are assumed to have the same importances. Practically, some criteria should have more influence than the others. To identify the most important criteria, it requires col-

lecting large amount empirical data and testing them with different models.

7.5 Future Work

In the future, further work can be done to extend the SERS or make a good use of it in the following aspects:

1. By collecting more empirical data of the commercial sites, it is possible to testify the correlation between the criteria. In total the SERS has defined 29 criteria, some of them have a closer correlation than the others. For example, if a system has a higher level of granularity (*I3*), it is very likely to have a smaller set of evaluators (*I2*). By analyzing the correlation among the criteria, it is able to identify the most important criteria of the model.
2. It is also possible to use the empirical data to conduct a more precise assessment on certain criteria. For instance, a sufficient number of evaluators is a key factor to influence whether the reputation information can reflect the true quality of the targets. With the help of empirical data, a mathematical model can be built to test the correlation between the targets and the ‘lowest’ sufficient number of reviews. Based on these analysis, an ideal system design can be built for each type of system.
3. The SERS is built for centralized reputation systems, which are mainly adopted by the commercial world. It can be extended to the distributed systems by revising the criteria and their quantifications and/or adding new criteria.
4. By taking a good use of the SERS, it is possible to predict the future trends of reputation systems. For example, the input collection cost (*I7*) has a great impact on the set of evaluators, which makes very few evaluators to provide rich media information. However, considering in the future, with the development of the information technologies, the cost of submitting rich media information might be hugely reduced; thus people may choose to use more pictures and videos as reputation information.

5. With the rising popularity of Mobile Internet Devices (MIDs), such as smart phones and tablet PCs, future research should pay attention to their influences on the reputation systems. Compared to traditional PCs and laptops, MIDs have different input and output interfaces. For example, many MIDs use touch screens instead of mice and keyboards for users to input data. These changes not only have impact on the design of reputation systems but also influence the collection costs of input and feedback loop.

Reputation system is a multi-disciplinary subject, covering economics, psychology, computer science and other relevant subjects. Although the SERS is built from the perspective of computer science, researchers from other areas can revise the model for their own interests.

References

- Aboody, D. and Lev, B. (2000). ‘Information asymmetry, r&d, and insider gains.’ *Journal of Finance*, **55**(6), 2747 – 2766.
- Akerlof, G. A. (1970). ‘The market for ‘lemons’: Quality uncertainty and the market mechanism.’ *The Quarterly Journal of Economics*, **84**(3), 488–500.
- Aladwani, A. M. and Palvia, P. C. (2002). ‘Developing and validating an instrument for measuring user-perceived web quality.’ *Information & Management*, **39**(6), 467 – 476.
- Aperjis, C. and Johari, R. (2010). ‘Optimal Windows for Aggregating Ratings in Electronic Marketplaces.’ *Management Science*, **56**(5), 864–880.
- Ba, S. and Pavlou, P. A. (2002). ‘Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior.’ *MIS Quarterly*, **26**(3), 243–268.
- Bailey, J. E. and Pearson, S. W. (1983). ‘Development of a Tool for Measuring and Analyzing Computer User Satisfaction.’ *Management Science*, **29**(5), 530–545.
- Ballou, D. P. and Pazer, H. L. (1985). ‘Modeling data and process quality in multi-input, multi-output information systems.’ *Management Science*, **31**(2), 150 – 162.
- Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009). ‘Methodologies for data quality assessment and improvement.’ *ACM Computing Surveys*, **41**(3), 1–52.
- Batini, C. and Scannapieco, M. (2006). *Data Quality — Concepts, Methodologies and Techniques*. Springer-Verlag Berlin Heidelberg.

- Bickart, B. and Schindler, R. M. (2001). 'Internet forums as influential sources of consumer information.' *Journal of Interactive Marketing*, **15**(3), 31.
- Bolton, G. E., Katok, E. and Ockenfels, A. (2004). 'How effective are electronic reputation mechanisms? an experimental investigation.' *Management Science*, **50**(11), 1587–1602.
- Bone, P. F. (1995). 'Word-of-mouth effects on short-term and long-term product judgments.' *Journal of Business Research*, **32**(3), 213–223.
- Bovee, M., Srivastava, R. P. and Mak, B. (2003). 'A conceptual framework and belief-function approach to assessing overall information quality.' *International Journal of Intelligent Systems*, **18**(1), 51–74.
- Buchegger, S., Mundinger, J. and Le Boudec, J.-Y. (2008). 'Reputation systems for self-organized networks.' *Technology and Society Magazine, IEEE*, **27**(1), 41–47.
- Chapman, A. D. (2005). *Principles of data quality*. Global Biodiversity Information Facility, Copenhagen.
- Chen, C. C. and Tseng, Y.-D. (2010). 'Quality evaluation of product reviews using an information quality framework.' *Decision Support Systems*, **50**(4), 755–768.
- Chen, J., Zhang, C. and Xu, Y. (2009). 'The role of mutual trust in building members' loyalty to a c2c platform provider.' *International Journal of Electronic Commerce*, **14**(1), 147–171.
- Chen, P., Wu, S. and Yoon, J. (2004). 'The impact of online recommendations and consumer feedback on sales.' In 'Proceedings of the International Conference on Information Systems,' 711–724.
- Chen, Y. and Xie, J. (2008). 'Online consumer review: Word-of-mouth as a new element of marketing communication mix.' *Management Science*, **54**(3), 477–491.
- Chen, Y.-H. and Barnes, S. (2007). 'Initial trust and online buyer behaviour.' *Industrial Management & Data Systems*, **107**(1), 21–36.

- Chevalier, J. A. and Mayzlin, D. (2006). 'The effect of word of mouth on sales: Online book reviews.' *Journal of Marketing Research (JMR)*, **43**(3), 345–354.
- Chewning, E. G. and Harrell, A. M. (1990). 'The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task.' *Accounting, Organizations and Society*, **15**(6), 527–542.
- Cho, J., Kwon, K. and Park, Y. (2009). 'Q-rater: A collaborative reputation system based on source credibility theory.' *Expert Systems with Applications*, **36**(2), 3751–3760.
- Clemons, E. K. (2008). 'How information changes consumer behavior and how consumer behavior determines corporate strategy.' *Journal of Management Information Systems*, **25**(2), 13 – 40.
- Cui, G., Lui, H.-k. and Guo, X. (2010). 'Online Reviews as a Driver of New Product Sales.' In '2010 International Conference on Management of e-Commerce and e-Government,' IEEE, 20–25.
- David, S. and Pinch, T. (2005). 'Six degrees of reputation: The use and abuse of online review and recommendation systems.' *Social Science Research Network Working Paper Series*.
- Davis, A. and Khazanchi, D. (2008). 'An empirical study of online word of mouth: as a predictor for multi-product category e-commerce sales.' *Electronic Markets*, **18**, 130–141.
- Davis, F. D. (1989). 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.' *MIS Quarterly*, **13**(3), 319.
- Dellarocas, C. (2001). 'Analyzing the economic efficiency of ebay-like online reputation reporting mechanisms.' In 'the 3rd ACM conference on Electronic Commerce (EC '01),' ACM Press, 171–179.
- Dellarocas, C. (2003). 'The digitization of word of mouth: Promise and challenges of online feedback mechanisms.' *Management Science*, **49**(10), 1407–1424.

- Dellarocas, C. (2005). 'Reputation mechanism design in online trading environments with pure moral hazard.' *Information Systems Research*, **16**(2), 209.
- Dellarocas, C. (2006a). 'How often should reputation mechanisms update a trader's reputation profile?' *Information Systems Research*, **17**(3), 345–354.
- Dellarocas, C. (2006b). 'Reputation mechanisms.' In 'Economics and Information Systems, Volume 1 (Handbooks in Information Systems)', , edited by T. Hendershott. Elsevier Science, 629–657.
- Dellarocas, C. (2009). 'The many faces of reputation: Towards a discipline of web 2.0 reputation system design.' Presented at the International Conference of Reputation – Theory and Technology.
- Dellarocas, C., Awad, N. and Zhang, X. M. (2004). 'Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning.' In 'the 25th International Conference on Information Systems (ICIS)', 34.
- DeLone, W. H. and McLean, E. R. (1992). 'Information systems success: The quest for the dependent variable.' *Information Systems Research*, **3**(1), 60–95.
- Delone, W. H. and McLean, E. R. (2003). 'The delone and mclean model of information systems success: A ten-year update.' *Journal of Management Information Systems*, **19**(4), 9–30.
- Delone, W. H. and Mclean, E. R. (2004). 'Measuring e-commerce success: Applying the delone & mclean information systems success model.' *International Journal of Electronic Commerce*, **9**(1), 31–47.
- Dewan, S. and Hsu, V. (2004). 'Adverse selection in electronic markets: Evidence from online stamp auctions.' *Journal of Industrial Economics*, **52**(4), 497–516.
- Dutton, W., di Gennaro, C. and Millwood Hargrave, A. (2005). 'Oxford internet survey 2005 report: The internet in britain.' Technical report, Oxford Internet Institute.

- Eppler, M. and Mengis, J. (2004). 'The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines.' *The Information Society*, **20**(5), 325–344.
- Falcone, R. and Castelfranchi, C. (2001). 'Social trust: A cognitive approach.' In 'Trust and deception in virtual societies,' , edited by C. Castelfranchi and Y.-H. Tan. Kluwer Academic Publishers, Norwell, MA, USA, 55–90.
- Fan, M., Tan, Y. and Whinston, A. B. (2005). 'Evaluation and design of online cooperative feedback mechanisms for reputation management.' *IEEE Transactions on Knowledge and Data Engineering*, **17**(2), 244–254.
- Farmer, F. R. and Glass, B. (2010). *Building web reputation systems*. O'Reilly and Yahoo! Press.
- Friedman, E., Resnick, P. and Sami, R. (2007). 'Manipulation-resistant reputation systems.' In 'Algorithmic Game Theory,' , edited by N. Nisan, T. Roughgarden, E. Tardos and V. V. Vazirani. Cambridge University Press, 677–697.
- Friedman, E. J. and Resnick, P. (2001). 'The social cost of cheap pseudonyms.' *Journal of Economics & Management Strategy*, **10**(2), 173–199.
- Frowe, I. (2005). 'Professional trust.' *British Journal of Educational Studies*, **53**(1), 34–53.
- Garcin, F., Faltings, B. and Jurca, R. (2009). 'Aggregating reputation feedback.' In 'the First International Conference on Reputation: Theory and Technology,' volume 1. 62–74.
- Gaur, V., Sharma, N. and Bedi, P. (2010). 'Evaluating Reputation Systems for Agent Mediated e-Commerce.' In 'Proceedings of International Conferences on advances in computer science 2010,' 220–224.
- Ghose, A., Ipeirotis, P. and Sundararajan, A. (2005). 'Reputation premiums in electronic peer-to-peer markets: Analyzing textual feedback and network structure.' In 'Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems,' ACM, 154.

- Godes, D. and Mayzlin, D. (2004). 'Using online conversations to study word-of-mouth communication.' *Marketing Science*, **23**(4), 545–560.
- Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. (1992). 'Using collaborative filtering to weave an information tapestry.' *Communications of the ACM*, **35**(12), 61–70.
- Good, D. (1988). 'Individuals, interpersonal relations, and trust.' In 'Trust: Making and Breaking Cooperative Relations,' , edited by D. Gambetta. Basil Blackwell, 31–48.
- Gregg, D. G. (2009). 'Outline reputation scores: How well are they understood?' *Journal of Computer Information Systems*, **50**(1), 90 – 97.
- Gregg, D. G. and Scott, J. E. (2006). 'The role of reputation systems in reducing on-line auction fraud.' *International Journal of Electronic Commerce*, **10**(3), 95–120.
- Gutowska, A. (2009). 'On Desideratum for B2C E-Commerce Reputation Systems.' *Journal of Computer Science and Technology*, **43**(12), 57–832.
- Hansen, J. P., Johansen, A. S., Hansen, D. W. and Itoh, K. (2003). 'Command without a click: Dwell time typing by mouse and gaze selections.' In 'Human-Computer Interaction - Interaction'03,' , edited by M. Rauterberg. IOS Press, 121 – 128.
- Healy, P. and Palepu, K. (2001). 'Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature.' *Journal of Accounting and Economics*, **31**(1-3), 405–440.
- Henneman, R. (1999). 'Design for usability: process, skills, and tools.' *Information, Knowledge, Systems Management*, **1**(2), 133–144.
- Hoffman, K., Zage, D. and Nita-Rotaru, C. (2009). 'A survey of attack and defense techniques for reputation systems.' *ACM Computing Surveys*, **42**(1), 1–31.

- Hoover, C. (2006). ‘A methodology for determining response time baselines.’ *In* ‘International CMG Conference,’ 85–94.
- Houser, D. and Wooders, J. (2006). ‘Reputation in auctions: Theory, and evidence from ebay.’ *Journal of Economics & Management Strategy*, **15**(2), 353–369.
- Hu, N., Liu, L. and Sambamurthy, V. (2010). ‘Fraud detection in online consumer reviews.’ *Decision Support Systems*, **50**(3), 614–626.
- Hu, N., Liu, L. and Zhang, J. J. (2008). ‘Do online reviews affect product sales? The role of reviewer characteristics and temporal effects.’ *Information Technology and Management*, **9**(3), 201–214.
- Hu, N., Pavlou, P. A. and Zhang, J. (2006). ‘Can online reviews reveal a product’s true quality?: empirical findings and analytical modeling of online word-of-mouth communication.’ *In* ‘Proceedings of the 7th ACM conference on Electronic commerce,’ EC ’06. ACM, 324–330.
- Huang, S., Shen, D., Feng, W., Baudin, C. and Zhang, Y. (2010). ‘Promote product reviews of high quality on e-commerce sites.’ *Pacific Asia Journal of the Association for Information Systems*, **2**(3), 51–71.
- Jarke, M., Jeusfeld, M. A., Quix, C. and Vassiliadis, P. (1999). ‘Architecture and quality in data warehouses: An extended repository approach.’ *Information Systems*, **24**(3), 229 – 253.
- Jin, G. Z. and Kato, A. (2006). ‘Price, quality and reputation: Evidence from an online field experiment.’ *RAND Journal of Economics*, **37**(4), 983–1005.
- Jindal, N. and Liu, B. (2007). ‘Review spam detection.’ *In* ‘Proceedings of the 16th international conference on World Wide Web,’ WWW ’07. 1189–1190.
- Jindal, N. and Liu, B. (2008). ‘Opinion spam and analysis.’ *In* ‘Proceedings of the international conference on Web search and web data mining,’ WSDM ’08. 219–230.

- Jøsang, A. and Ismail, R. (2002). ‘The beta reputation system.’ *In* ‘the 15th Bled Conference on Electronic Commerce,’ 324–337.
- Jøsang, A., Ismail, R. and Boyd, C. (2007). ‘A survey of trust and reputation systems for online service provision.’ *Decision Support Systems*, **43**(2), 618–644.
- Jurca, R. (2008). *Truthful reputation mechanisms for online systems*. Ph.D. thesis, Ecole Polytechnique Fédéral de Lausanne (EPFL).
- Kahn, B. K., Strong, D. M. and Wang, R. Y. (2002). ‘Information quality benchmarks: product and service performance.’ *Communications of the ACM*, **45**(4), 184–192.
- Karat, C.-M., Halverson, C., Horn, D. and Karat, J. (1999). ‘Patterns of entry and correction in large vocabulary continuous speech recognition systems.’ *In* ‘CHI ’99: the SIGCHI conference on Human factors in computing systems,’ ACM Press, 568–575.
- Katerattanakul, P. and Siau, K. (1999). ‘Measuring information quality of web sites: development of an instrument.’ *In* ‘the 20th International Conference on Information Systems,’ ICIS ’99. Association for Information Systems, 279–285.
- Klang, M. (2001). ‘Who do you trust? beyond encryption, secure e-business.’ *Decision Support Systems*, **31**(3), 293–301.
- Koh, N. S., Hu, N. and Clemons, E. K. (2010). ‘Do online reviews reflect a product’s true perceived quality? An investigation of online movie reviews across cultures.’ *Electronic Commerce Research and Applications*, **9**(5), 374–385.
- Koufaris, M. and Hampton-Sosa, W. (2004). ‘The development of initial trust in an online company by new customers.’ *Information & Management*, **41**(3), 377–397.
- Kramer, R. M. (1999). ‘Trust and distrust in organizations: Emerging perspectives, enduring questions.’ *Annual Review of Psychology*, **50**(1), 569–598.

- Kumar, N. and Benbasat, I. (2006). 'The influence of recommendations and consumer reviews on evaluations of websites.' *Information Systems Research*, **17**(4), 425–439.
- Laat, P. (2005). 'Trusting virtual trust.' *Ethics and Information Technology*, **7**(3), 167–180.
- Lampe, C. and Resnick, P. (2004). 'Slash(dot) and burn: distributed moderation in a large online conversation space.' In 'CHI '04: the SIGCHI conference on Human factors in computing systems,' 543–550.
- Lau, R. Y., Liao, S. S. and Xu, K. (2010). 'An Empirical Study of Online Consumer Review Spam: A Design Science Approach.' In 'ICIS 2010 Proceedings,' 1–21.
- Laudon, J. and Laudon, K. (2007). *Management Information Systems: Managing the Digital Firm & Multimedia*. Prentice Hall, 10th edition.
- Lerman, K. (2006). 'Social networks and social information filtering on digg.' unpublished.
- Lerman, K. and Galstyan, A. (2008). 'Analysis of social voting patterns on digg.' In 'Proceedings of the first workshop on Online social networks,' WOSP '08. 7–12.
- Levitin, A. (2001). *Introduction to the design and analysis of algorithms*. Addison Wesley, 2nd edition.
- Lewis-beck, M. S. (ed.) (1993). *Basic Statistics*. SAGE Publications.
- Li, X. and Hitt, L. M. (2008). 'Self-selection and information role of online product reviews.' *Information Systems Research*, **19**(4), 456–474.
- Li, X. and Hitt, L. M. (2010). 'Price effects in online product reviews: An analytical model and empirical analysis.' *MIS Quarterly*, **34**(4), 809–A5.
- Liang, Z. and Shi, W. (2005). 'Performance evaluation of rating aggregation algorithms in reputation systems.' In 'International Conference on Collaborative Computing: Networking, Applications and Worksharing,' IEEE Computer Society, 10.

- Liu, C. and Arnett, K. P. (2000). 'Exploring the factors associated with web site success in the context of electronic commerce.' *Information & Management*, **38**(1), 23 – 33.
- Liu, L. and Chi, L. (2002). 'Evolutionary data quality: A theory-specific view.' *In* '7th International Conference on Information Quality (ICIQ 2002),' 292–304.
- Liu, Y. (2006). 'Word-of-mouth for movies: Its dynamics and impact on box office revenue.' *Journal of Marketing*, **70**(3), 74–89.
- Longstreet, P. (2010). 'Evaluating Website Quality: Applying Cue Utilization Theory to WebQual.' *In* '43rd Hawaii International Conference System Sciences (HICSS),' IEEE Computer Society, 1–7.
- Lucking-Reiley, D., Bryan, D., Prasad, N. and Reeves, D. (2007). 'Pennies from ebay: The determinants of price in online auctions.' *Journal of Industrial Economics*, **55**(2), 223–233.
- Malaga, R. A. (2001). 'Web-based reputation management systems: Problems and suggested solutions.' *Electronic Commerce Research*, **1**(4), 403–417.
- McDonald, C. G. and Slawson, V. C. (2002). 'Reputation in an internet auction market.' *Economic Inquiry*, **40**(4), 633–650.
- McKinney, V., Yoon, K. and ?Mariam? Zahedi, F. (2002). 'The measurement of web-customer satisfaction: An expectation and disconfirmation approach.' *Information Systems Research*, **13**(3), 296–315.
- Metzger, M. J. (2007). 'Making sense of credibility on the web: Models for evaluating online information and recommendations for future research.' *Journal of the American Society for Information Science and Technology*, **58**(13), 2078–2091.
- Miller, R. B. (1968). 'Response time in man-computer conversational transactions.' *In* '1968 Proceedings of the Fall Joint Computer Conference,' ACM, 267–277.
- Misztal, B. A. (1996). *Trust in Modern Societies*. Polity Press.

- Molla, A. and Licker, P. (2001). 'E-commerce systems success: An attempt to extend and respecify the delone and maclean model of is success.' *Journal of Electronic Commerce Research*, **2**(4), 131–141.
- Morzy, M. (2008). 'New algorithms for mining the reputation of participants of online auctions.' *Algorithmica*, **52**(1), 95–112.
- Mudambi, S. M. and Schuff, D. (2010). 'What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com.' *MIS Quarterly*, **34**(1), 185–200.
- Mui, L., Mohtashemi, M. and Halberstadt, A. (2002a). 'A computational model of trust and reputation.' In 'the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 7,' 188.
- Mui, L., Mohtashemi, M. and Halberstadt, A. (2002b). 'Notions of reputation in multi-agents systems: a review.' In 'AAMAS '02: the first international joint conference on Autonomous agents and multiagent systems,' ACM, 280–287.
- Neill, M. O., Wright, C. and Fitz, F. (2001). 'Quality evaluation in environments : an application of the performance measurement technique.' *Managing Service Quality*, **11**(6), 402–417.
- Nielsen (2008). 'Nielsen online holiday survey 2008.' Online Survey.
- Nissenbaum, H. (2001). 'Securing trust online: Wisdom or oxymoron?' *Boston University Law Review*, **81**, 635–664.
- Novak, T. P., Hoffman, D. L. and Yung, Y.-F. (2000). 'Measuring the customer experience in online environments: A structural modeling approach.' *Marketing Science*, **19**(1), 22–42.
- O'Reilly III, C. A. (1980). 'Individuals and information overload in organizations: Is more necessarily better?.' *Academy of Management Journal*, **23**(4), 684–696.
- Palmer, J. W. (2002). 'Web site usability, design, and performance metrics.' *Information Systems Research*, **13**(2), 151–167.

- Parasuraman, A., Zeithaml, V. and Berry, L. (2004). 'Refinement and reassessment of the SERVQUAL scale.' *Journal of Retailing*, **67**(4), 114.
- Pather, S. and Usabuwera, S. (2010). 'Implications of e-Service Quality Dimensions for the Information Systems Function.' In '43rd Hawaii International Conference on System Sciences (HICSS),' IEEE, 1–10.
- Pavlou, P. A. and Dimoka, A. (2006). 'The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation.' *Information Systems Research*, **17**(4), 392–414.
- Pernici, B. and Scannapieco, M. (2002). 'Data quality in web information systems.' In 'the 21st International Conference on Conceptual Modeling,' 397–413.
- Petter, S., Delone, W. and Mclean, E. (2008). 'Measuring information systems success: models, dimensions, measures, and interrelationships.' *European Journal of Information Systems*, **17**(3), 236–263.
- Pettit, P. (2004). 'Trust, reliance and the internet.' *Analyse und Kritik*, **26**, 108–121.
- Pierce, J. (2008). 'The world internet project report 2009.' Technical report, The World Internet Project.
- Pindyck, R. S. and Rubinfeld, D. L. (2009). *Microeconomics*. Pearson Education, Inc, 7th edition.
- Pitt, L. F., Watson, R. T. and Kavan, C. B. (1995). 'Service Quality: A Measure of Information Systems Effectiveness.' *MIS Quarterly*, **19**(2), 173.
- Rasmusen, E. (2001). *Games & Information — An Introduction to Game Theory*. Blackwell Publishers Ltd, 108 Cowley Road, Oxford, 3rd edition.
- Redman, T. C. (1997). *Data Quality for the Information Age*. Artech House, Inc., Norwood, MA, USA.
- Resnick, P., Kuwabara, K., Zeckhauser, R. and Friedman, E. (2000). 'Reputation systems.' *Communications of the ACM*, **43**(12), 45–48.

- Resnick, P., Zeckhauser, R., Swanson, J. and Lockwood, K. (2006). ‘The value of reputation on ebay: A controlled experiment.’ *Experimental Economics*, **9**(2), 79–101.
- Rubicon Consulting (2008). ‘Online communities and their impact on business: Ignore at your peril.’ White papers, Rubicon Consulting.
- Ruohomaa, S., Kutvonen, L. and Koutrouli, E. (2007). ‘Reputation management survey.’ In ‘ARES ’07: The Second International Conference on Availability, Reliability and Security,’ 103–111.
- Sabater, J. and Sierra, C. (2001). ‘Regret: reputation in gregarious societies.’ In ‘AGENTS ’01: the fifth international conference on Autonomous agents,’ 194–195.
- Sabater, J. and Sierra, C. (2005). ‘Review on computational trust and reputation models.’ *Artificial Intelligence Review*, **24**(1), 33–60.
- Scannapieco, M. and Catarci, T. (2002). ‘Data quality under the computer science perspective.’ *Archivi & Computer*, **2**.
- Seddon, P. B. (1997). ‘A respecification and extension of the delone and mclean model of is success.’ *Information Systems Research*, **8**(3), 240–253.
- Shannon, C. E. (1951). ‘Prediction and entropy of printed english.’ *Bell Systems Technical Journal*, **30**, 50–64.
- Simons, H. (2009). *Case Study Research in Practice*. SAGE Publications Inc.
- Smith, A. G. (1997). ‘Testing the surf: Criteria for evaluating internet information resources.’ *The Public-Access Computer Systems Review*, **8**(3).
- Spiller, P. and Lohse, G. (1997). ‘A classification of Internet retail stores.’ *International Journal of Electronic Commerce*, **2**(2), 29–56.
- Stair, R. M., Reynolds, G. and Reynolds, G. W. (2010). *Principles of Information System*. Course Technology, 9th edition.

- Straub, D. and Carlson, C. L. (1989). 'Validating instruments in MIS research.' *MIS quarterly*, **13**(2), 147–169.
- Surowiecki, J. (2005). *The Wisdom of Crowds*. Anchor.
- Swamynathan, G., Almeroth, K. C. and Zhao, B. Y. (2010). 'The design of a reliable reputation system.' *Electronic Commerce Research*, **10**(3-4), 239–270.
- Tang, L. C., Zhao, Y., Austin, S., Darlington, M. and Culley, S. (2008). 'A characteristic based information evaluation model.' In 'the 2nd ACM workshop on Information credibility on the web (WICOW '08),' 89–92.
- Tate, M. and Evermann, J. (2009). 'Descendents of ServQual in Online Services Research: The End of the Line.' In 'the 15th Americas Conference on Information Systems (AMCIS),' 1–11.
- The Oxford English Dictionary (1992). *The Oxford English Dictionary*. Clarendon Press, 2nd edition.
- Tran, N., Min, B., Li, J. and Subramanian, L. (2009). 'Sybil-resilient online content voting.' In 'Proceedings of the 6th USENIX symposium on Networked systems design and implementation,' Berkeley, CA, USA, 15–28.
- Wang, R. W. and Strong, D. M. (1996). 'Beyond accuracy: What data quality means to data consumers.' *Journal of Management Information Systems*, **12**(4), 5–33.
- Wang, Y. D. and Emurian, H. H. (2005). 'An overview of online trust: Concepts, elements, and implications.' *Computers in Human Behavior*, **21**(1), 105–125.
- Weinreich, H., Obendorf, H., Herder, E. and Mayer, M. (2008). 'Not quite the average: An empirical study of web use.' *ACM Transactions on the Web*, **2**(1), 1–31.
- Westbrook, R. A. (1987). 'Product/Consumption-Based Affective Responses and Postpurchase Processes.' *Journal of Marketing Research*, **24**(3), 258–270.

- Whitby, A., Jøsang, A. and Indulska, J. (2004). 'Filtering out unfair ratings in bayesian reputation systems.' *In* 'the 7th International Workshop on Trust in Agent Societies,' 106–117.
- Whitmeyer, J. M. (2000). 'Effects of positive reputation systems.' *Social Science Research*, **29**(2), 188–207.
- Wilson, R. (1985). 'Reputations in games and markets.' *In* 'Game-theoretic models of bargaining,' , edited by A. E. Roth. Cambridge University Press, 27–62.
- Yang, J., Hu, X. and Zhang, H. (2007). 'Effects of a reputation feedback system on an online consumer-to-consumer auction market.' *Decision Support Systems*, **44**(1), 93–105.
- Yang, Z., Peterson, R. T. and Cai, S. (2003). 'Services quality dimensions of Internet retailing: an exploratory analysis.' *Journal of Services Marketing*, **17**(7), 685–700.
- Yin, R. K. (2003). *Applications of Case Study Research*. SAGE Publications Inc, 2nd edition.
- Yin, R. K. (2009). *Case Study Research — Design and Mehtods*. SAGE Publications Inc, 4th edition.
- Yoon, C. and Kim, S. (2009). 'Developing the Causal Model of Online Store Success.' *Journal of Organizational Computing and Electronic Commerce*, **19**(4), 265–284.
- Zacharia, G. and Maes, P. (2000). 'Trust management through reputation mechanisms.' *Applied Artificial Intelligence*, **14**(9), 881–907.
- Zhang, R. and Tran, T. (2011). 'An information gain-based approach for recommending useful product reviews.' *Knowledge and Information Systems*, **26**, 419–434.
- Zhang, X. and Dellarocas, C. (2006). 'The Lord of the Ratings: How a Movies Fate is Influenced by Reviews?' *In* 'Proceedings of the 27th International Conference on Information Systems (ICIS 2006),' 1959–1978.

- Zhao, Y., Tang, L., Darlington, M., Austin, S. and Culley, S. (2008). 'High value information in engineering organisations.' *International Journal of Information Management*, **28**(4), 246 – 258.
- Zheng, W. and Jin, L. (2009). 'Online reputation systems in web 2.0 era.' In 'Americas Conference on Information Systems (AMCIS) Proceedings,' 296–306.
- Zhou, M., Dresner, M. and Windle, R. (2009). 'Revisiting feedback systems: Trust building in digital markets.' *Information & Management*, **46**(5), 279–284.
- Zhu, F. and Zhang, X. M. (2010). 'Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics.' *Journal of Marketing*, **74**(2), 133–148.
- Zhu, Y. (2010). 'Measurement and analysis of an online content voting network: a case study of digg.' In 'Proceedings of the 19th international conference on World wide web,' WWW '10. 1039–1048.